CHANNEL ESTIMATION IN MULTIPLE-INPUT MULTIPLE-OUTPUT SYSTEMS

By

BEOMJIN PARK

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2004

Copyright 2004

by

Beomjin Park

To my parents, Kyung-Ho Park and Bong-Hee Kong.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Dr. Tan F. Wong, for his energetic and passionate guidance throughout my Ph.D. program. Without his continuous and patient guidance, this work never would have been accomplished.

I would also like to thank Dr. Yuguang "Michael" Fang, Dr. John M. Shea, and Dr. Louis N. Cattafesta III for their supporting roles as my committee members.

Last but not least, I would like to thank my family for always encouraging and trusting me throughout my whole life. Their endless support encouraged me to finish this valuable accomplishment.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

ABSTRACT x

CHAPTER

1 INTRODUCTION 1

1.1 Previous Work 3
1.2 Problem Approach 6
1.3 Organization of This Dissertation 7

2 BACKGROUND 9

2.1 Space-Time Wireless Communication Systems 9
2.1.1 Space-Time Signal Model 9
2.1.2 MIMO Channel Model 10
2.2 Minimum Variance Unbiased Estimation 14
2.2.1 Unbiased Estimator 14
2.2.2 Minimum Variance Criterion 15
2.2.3 Linear Models 15
2.2.4 Best Linear Unbiased Estimator 17
2.2.5 Maximum Likelihood Estimator 20
2.3 Bayesian Estimator 21
2.4 Circulant Matrices and Toeplitz Matrices 23
2.4.1 Circulant Matrices 24
2.4.2 Toeplitz Matrices 25
2.4.3 Absolutely Summable Toeplitz Matrix 26
2.5 Summary 29

3 SYSTEM MODEL AND CHANNEL ESTIMATION 30

3.1 Channel Estimation with BLUE 30
3.1.1 System Model 30
3.1.2 Best Linear Unbiased Channel Estimator 32
3.2 Channel Estimation with Bayesian Estimator 38
3.2.1 System Model 38

3.2.2 Bayesian Channel Estimator 39
3.3 Summary 43
3.4 Derivation of the **J** Matrix 43

4 TRAINING SEQUENCE OPTIMIZATION 48

4.1 Optimal Training Sequence Set with BLUE 48
4.2 Optimal Training Sequence Set with Bayesian Estimator 51
4.3 Conclusion 54
4.4 Solution of the Optimization Problem 54

5 FEEDBACK DESIGN 58

5.1 Feedback Design for BLUE 58
5.2 Feedback Design for Bayesian Estimator 60
5.3 Summary 63

6 NUMERICAL RESULTS 64

6.1 Asymptotic Estimation Performance Gain 64
6.1.1 BLUE 64
6.1.2 Bayesian Estimatior 66
6.2 Numerical Examples for BLUE 67
6.2.1 AR(1) Jammer 68
6.2.2 Co-channel Interferer 72
6.2.3 Average MSE with Hadamard Sequence Set 77
6.3 Numerical Examples for Bayesian Estimator 83
6.3.1 AR(1) Jammer 83
6.3.2 Co-channel interferer 84
6.3.3 Bit Error Rate Performance 92
6.4 Conclusion 92

7 CONCLUSION AND FUTURE WORK 96

7.1 Conclusion 96
7.2 Future Work 97

REFERENCES 100

BIOGRAPHICAL SKETCH 103

LIST OF TABLES

Table page

3–1 Matrix Notation 47

6–1 Comparison of asymptotic maximum MSE reduction ratio and MSE ratio between using optimal and Hadamard sequences in the case of AR jammers. 95

6–2 Comparison of asymptotic maximum MSE reduction ratio and MSE ratio between using optimal and Hadamard sequences in the case of co-channel interferers. 95

LIST OF FIGURES

Figure page

2–1 MIMO channel consisting of n_t transmit n_r receive antennas. 10

2–2 Block transmission [24]. 13

2–3 Definition of time index for the block transmission [24]. 13

6–1 Comparison of MSEs obtained by using different training sequence sets. AR-1 jammer with $\alpha = 0.7$ and $n_t = n_r = 2$. 70

6–2 Comparison of MSEs obtained by using different training sequence sets. AR-1 jammer with $\alpha = 0.9$ and $n_t = n_r = 2$. 71

6–3 Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with rectangular waveform, $\tau = 0.3T$, and $n_t = n_r = 2$. 75

6–4 Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with rectangular waveform, $\tau = 0.5T$, and $n_t = n_r = 2$. 76

6–5 Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with raised-cosine waveform, $\beta = 0.5$, $\tau = 0.3T$, and $n_t = n_r = 2$. 78

6–6 Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with raised-cosine waveform, $\beta = 0.5$, $\tau = 0.5T$, and $n_t = n_r = 2$. 79

6–7 Comparison MSEs obtained by using optimal training sequence set with average MSEs obtained by using all possible Hadamard training sequence set. AR(1) jammer with $\alpha = 0.7$ and $n_t = n_r = 2$. 80

6–8 Comparison MSEs obtained by using optimal training sequence set with average MSEs obtained by using all possible Hadamard training sequence set. AR(1) jammer with $\alpha = 0.9$ and $n_t = n_r = 2$. 81

6–9 Comparison MSEs obtained by using optimal training sequence set with average MSEs obtained by using all possible Hadamard training sequence set. Co-channel interferer with rectangular waveform, $\tau = 0.3T$ and $n_t = n_r = 2$. 82

6–10 Comparison of MSEs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.3$ and $\alpha_2 = 0.5$. 85

6–11 Comparison of MSEs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$. 86

6–12 Comparison of MSEs obtained by using different training sequence sets. Two co-channel interferers with rectangular waveforms and delays $\tau_1 = 0.3T$, $\tau_2 = 0.5T$. 89

6–13 Comparison of MSEs obtained by using different training sequence sets. Two co-channel interferers with ISI-free waveform waveforms and delays $\tau_1 = 0.3T$, $\tau_2 = 0.5T$. 91

6–14 Comparison of BERs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$. 93

6–15 Comparison of BERs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$. 94

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CHANNEL ESTIMATION IN MULTIPLE-INPUT MULTIPLE-OUTPUT SYSTEMS

By

Beomjin Park

May 2004

Chair: Tan F. Wong
Major Department: Electrical and Computer Engineering

We address the problems of channel estimation and optimal training sequence design for multiple-input and multiple-output (MIMO) systems over flat fading channels in the presence of colored interference. In practice, information of the unknown channel parameters is often obtained by sending known training symbols to the receiver. During the training period, we obtain the estimates of the channel parameters based on the received training block. This method is called training based channel estimation. In order to estimate unknown channel parameters, we employ two different channel estimators - the best linear unbiased estimator (BLUE) and Bayesian channel estimator. We consider the BLUE for the case where there is a single interferer with the deterministic channel assumption. We consider the Bayesian channel estimator for the case where there are multiple interferers with the assumption of random channels. We note that the mean square error (MSEs) of the channel estimators are dependent on the choice of the training sequence set. Hence we determine the optimal training sequence set that can minimize the MSEs of the channel estimators under a total transmit power constraint. In order to obtain the advantage of

the optimal training sequence design, long-term statistics of the interference correlation are needed at the transmitter. Hence this information needs to be estimated at the receiver and fed back to the transmitter. It is desirable that if we can reduce the estimation error of the short-term channel fading parameters by using a minimal amount of information that is fed back from the receiver. We develop such a feedback strategy to design an approximate optimal training sequence set in this work.

CHAPTER 1
INTRODUCTION

With the emergence of next-generation wireless mobile communications, there is an increasing demand for higher data rates, better quality of service, and higher network capacity. In an effort to support such demand within the limited availability of radio frequency spectrum, many researchers have begun to utilize not only the time and frequency dimensions but also the space dimension to design communication systems with higher spectral efficiencies. Recent research in information theory has shown that large gains in reliability of communications over wireless channels can be achieved by exploiting spatial diversity [1, 2]. The concept of spatial diversity is that, in the presence of random fading caused by multi-path propagation, the signal-to-noise ratio (SNR) can be significantly improved by combining the outputs of decorrelated antenna elements. Such space utilization can be usually obtained by using multiple antenna elements arranged in an array at both the transmitter and receiver. Furthermore, it has been reported that multiple antennas along with space-time coding (STC) or diversity techniques can aggressively exploit multi-path propagation effects for the benefit of improving the communication capability of a system. Recently, wireless communication systems using multiple antennas, usually referred to as multiple-input multiple-output (MIMO) systems, have drawn considerable attention, because MIMO systems promise higher capacity [1, 2] than single-antenna systems over fading channels. As described in Gesbert et al. [3], the idea behind MIMO is that the signals on the transmit antennas at one end and the receive antennas at the other end are "combined" in such a way that the quality (bit-error rate or BER) or the data rate (bits/sec) of communication for each MIMO user will be improved. Different STC techniques [4-8] have been proposed to practically achieve

the capacity advantages of MIMO systems. STC is a set of practical signal design techniques aimed at approaching the information theoretic capacity limit of MIMO channels. The fundamentals of STC were established by Tarokh et al. in 1998 [4]. Among STC techniques, the main classes are Bell Labs layered space-time (BLAST) architecture proposed by Foschini [1], space-time trellis codes (STTC) proposed by Tarokh et al. [4], and space-time block codes (STBC) proposed by Alamouti [8].

Moreover major impairments such as fading, delay spread, and co-channel interference caused by the wireless communication channels can be further mitigated by employing MIMO systems. In a multi-path fading environment, the transmitted signal is scattered by objects such as buildings, trees, or mountains before reaching the receiver. This causes the signal to fade. While this scattering is detrimental to conventional wireless transmission, MIMO systems use multi-path propagation to increase the data transmission rate due to the spatial diversity. Spatial diversity can be achieved by sufficiently spaced multiple antennas at the receiver so that multiple copies of transmitted signal propagated through channels with different fading are obtained. Because there exists only a small probability that all signal copies are in a deep fade simultaneously, spatial diversity can increase robustness of the wireless link and can be used to obtain higher data throughput. Interference suppression can be achieved by using the spatial dimension provided by multiple antenna elements in MIMO systems. Hence the system is less susceptible to interference. This also can lead to system capacity improvement.

With the many advantages mentioned above, MIMO system designs begun to be applied in commercial wireless products and networks such as broadband wireless access systems, wireless local area networks (WLAN), and third generation (3G) networks. MIMO systems with sophisticated space-time processing techniques could be the next frontier in wireless communications and we could see many other applications in near future.

1.1 Previous Work

Much work has been done to design training sequences for channel estimation in single-antenna systems [9-14]. In Crozier et al. [9], a least sum of squared errors (LSSE) channel estimation algorithm that is used to estimate the initial channel response from a short preamble training sequence was presented. To determine the quality of training sequence for a given channel, normalized signal-to-estimation-error ratio (SER), normalized with respect to SNR, is used. A method of generating "perfect" preamble training sequences, whose associated preamble correlation matrix is perfectly diagonal so that mean squared channel estimation error is minimized, was introduced. In addition, it was shown that perfect training sequences can always be obtained for any given channel response length. A computer search was performed to find the best preamble sequences for given numbers of channel taps and preamble lengths. In Mow [15, 16], the perfect root-of-unity sequences (PRUS) were proposed for different applications. A root-of-unity sequence is a sequence whose elements are all complex roots of unity in the form of $\exp(j2\pi r)$, with r a rational number, where $j=\sqrt{-1}$ [16]. The construction method of complex codes of the form $\exp(j\alpha)$ with good periodic correlation properties without the restriction of code length was proposed in Chu [17]. This code is called the polyphase code.

Training sequence design for the block adaptive channel estimation method for direct-sequence/code-division multiple access (DS/CDMA) was considered in Caire and Mitra [10]. A minimum-mean squared error (MMSE) channel estimator was used and its normalized estimation mean squared error (MSE) was obtained. Optimal training sequences were designed through minimization of the resulting MSE. As a result, Caire and Mitra obtained an optimal set of training sequences that must satisfy $\mathbf{S}^H\mathbf{S} = \epsilon\mathbf{I}$, where $\mathbf{S}$ is the training symbol matrix with block circulant structure, ϵ is proportional to the common energy of training sequences, and $\mathbf{I}$ is an identity matrix

[10]. Caire and Mitra constructed optimal training sequences by using the set of root-of-unity sequences [15, 16] that satisfy this requirement.

In Tellambura et al. [11], the least-square estimates of the channel impulse response obtained by using a known aperiodic sequence was considered. In addition, Tellambura et al. described how to find optimum aperiodic sequences so that it offers the best possible signal-to-estimation-error ratio (SER) at the output of the channel estimator. A performance measure was proposed to assess the quality of a binary sequence for channel estimation by using the trace of the inverse of its associated autocorrelation matrix.

Tellambura et al. [12] discussed the problem of selecting the optimum training sequence for channel estimation in frequency domain over a time-dispersive channel by using discrete Fourier transform. Tellambura et al. introduced a search criterion, termed the gain loss factor (GLF), which minimizes the variance of the estimation error. Theoretical upper and lower bounds on the GLF were derived. Moreover, an optimal sequence search procedure for periodic and aperiodic cases was provided. However the sequences obtained by computer search in this work were optimal only for frequency domain.

Chen and Mitra [13] compared the frequency-domain training sequence optimization technique introduced by Tellambura et al. [12] and the time-domain channel estimation method introduced by Crozier et al. [9]. Chen and Mitra employed the GLF defined in Tellambura et al. [12] to compare the time-domain method to the frequency-domain method. The results showed that the time-domain method achieves a smaller mean-squared channel estimation error over the frequency-domain technique with a significantly higher optimal training sequence search complexity. In addition, Chen and Mitra proposed an alternative search criterion that can provide equivalent or better performance than the frequency-domain method with a lower search complexity.

An information-theoretic approach for finding the optimal amount of training for frequency-selective channels was introduced by Vikalo et al. [14]. By using a lower bound on the channel capacity of a training based transmission scheme, Vikalo et al. determined the optimal training parameters that maximize this lower bound. These parameters are the length of the training interval, training data sequence, and training power. Vikalo et al. showed that the optimal number of training symbols is equal to the length of the channel impulse response.

Channel estimation in multiple-antenna systems has also been considered [18-21]. In particular, the channel estimation problem for MIMO systems over flat fading channels was considered in Marzetta [18] and Naguib et al. [19]. In Marzetta [18], Marzetta obtained the optimum training signal that minimizes the covariance of the maximum likelihood (ML) estimator under a total energy constraint. Marzetta claimed that the duration of the training interval must be at least as large as the number of transmit antennas. In Hassibi and Hochwald [20], the problem of determining the optimal number of training symbols in MIMO systems over flat fading channels was addressed. This was an extended work based on Vikalo [14]. In Fragouli et al. [21], methods were proposed to reduce the complexity of designing training sequences over frequency-selective channels in MIMO systems.

An information-theoretic approach has been used to optimize the design of the training scheme over frequency-selective channels [23]. Ma et al. obtained optimal training parameters that can maximize a lower bound on the average channel capacity. They showed that this approach is the minimization of the MMSE channel estimation error.

In summary, there have been two major approaches to design optimal training sequences for both single antenna systems [9-14] and multiple antenna systems [18-23]. One approach is to find training sequences that minimize the channel estimation

error [9-13], [18], [21] and other approach is to maximize lower bounds of the channel capacity [14], [20], [23]. Most of the above cited works consider estimation of the channel parameters in the presence of white noise. The general result reported in these cases is to find training sequences whose associated correlation matrix is perfectly diagonal, i.e., a scalar times the identity matrix. Based on this observation, aperiodic, periodic, maximal-length sequences (m-sequences), and perfect root of unity sequences have been used to construct optimal training sequences.

1.2 Problem Approach

As indicated in Section 1.1, much work has been done for finding optimal training sequences with a white noise model for both single antenna and multiple antenna systems. In this work, we address the problem of training sequence design for MIMO systems over flat fading channels in the presence of colored interference. The colored interference model is more suitable than the white noise model when jammers and co-channel interferers are present in the system.

To be able to achieve coding advantage provided by space-time coding schemes and the other advantages by MIMO systems mentioned above, it is required to obtain accurate channel information at the receiver. In this work, we employ the training based channel estimation approach to do so. During the training period, we obtain the information of the channel parameters by employing two different channel estimators - the best linear unbiased estimator (BLUE) and the Bayesian channel estimator. We consider two different MIMO system models. We assume that there is a single interferer in the BLUE approach. The multiple interferer case is considered under the Bayesian approach. The major difference between these two approach is that the channel is assumed to be unknown but deterministic in the BLUE approach, while the channel is assumed to be random with a known distribution in the Bayesian approach. The mean squared errors (MSE) of the two channel estimators are used as performance metrics for selecting the training sequence set.

In the BLUE approach, we show that when the interference covariance matrix decomposes into a Kronecker product of temporal and spatial correlation matrices, only the temporal correlation needs to be considered in obtaining the optimal training sequence set. Then, we determine the optimal training sequence set that minimizes the MSE under a total transmit power constraint. In the Bayesian approach, we show that the MSE of the channel estimator depends on the choice of training symbol matrix without any restriction on the structure of the interference covariance matrix. We also select the optimal training sequence set that minimizes the MSE of the estimator with the total energy constraint in this case.

In order to obtain the advantage of the optimal training sequence design, long-term statistics of the interference correlation are needed at the transmitter. Hence this information needs to be estimated at the receiver and fed back to the transmitter. Obviously it is desirable that only a minimal amount of information is needed to be fed back from the receiver to gain the advantage in reducing the estimation error of the short-term channel fading matrix. We develop an information feedback scheme that requires a minimal amount of information to be fed back from the receiver to approximately obtain the optimal training sequence set at the transmitter.

Numerical results show that we can reduce the MSEs of the channel estimators significantly by using the optimal training sequence set instead of a usual orthogonal training sequence set. We can also obtain comparable performance with the approximate optimal training sequence set obtained by the proposed feedback scheme.

1.3 Organization of This Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, we briefly summarize some background estimation and matrix analysis results that will be used in dissertation. Channel models for the MIMO systems are introduced in Section 2.1.2. In Section 2.2, we address the minimum unbiased estimator with minimum variance criterion. The best linear unbiased estimator (BLUE) and maximum likelihood (ML)

estimator are introduced. The Bayesian estimator is discussed in Section 2.3. In addition, results regarding circulant and Toeplitz matrices are introduced in Section 2.4. We also discuss asymptotic behaviors of these two types of matrices. We describe the MIMO system model and develop the BLUE and Bayesian channel estimator for the channel matrix based on the received training sequence block in Chapter 3. The MSEs, which will be used as a performance metric, for both estimators are obtained. In Section 3.1, we consider both the case of non-singular and singular interference covariance matrix in the BLUE approach. In Section 3.2, channel estimation with the Bayesian estimator is discussed. In this section, we consider the case where there exist multiple interferers. In Chapter 4, the training sequence optimization problem is considered and its optimal solution is given. In Chapter 5, we develop the feedback scheme to approximately obtain the optimal training sequence set. Numerical results are given in Chapter 6. The MSEs of the BLUE and Bayesian estimators are given and compared by using different training sequence sets. In Chapter 7, conclusions and future work are addressed.

CHAPTER 2
BACKGROUND

In this chapter, we introduce space-time wireless communication systems with antenna arrays at both the transmitter and receiver. Channel models for wireless communication systems using multiple antennas, referred to as multiple-input multiple-output (MIMO) systems, are discussed in this chapter. We also discuss minimum variance unbiased (MVU) estimation of the unknown channel parameters. We briefly introduce linear unbiased estimators such as the best linear unbiased estimator (BLUE) and maximum likelihood (ML) estimator. In addition, we discuss the Bayesian estimator, which is also know as minimum mean square error (MMSE) estimator, and its properties. Finally, we discuss the properties of circulant matrices and Toeplitz matrices. In addition, we study asymptotic behavior of these two types of matrices.

2.1 Space-Time Wireless Communication Systems

2.1.1 Space-Time Signal Model

Different space-time wireless communication systems consisting of transmitter, radio channel, and receiver can be categorized by the numbers of inputs and outputs. The conventional configuration is to have a single antenna at each side of the radio channel; hence a single-input single-output (SISO) system results. In a similar manner, multiple-input single-output (MISO) system, a single-input multiple-output (SIMO) system, and a multiple-input multiple-output (MIMO) system would result when the system has multiple antennas at the transmitter, multiple antennas at the receiver, and multiple antennas at both the transmitter and receiver, respectively. Thus we can consider SISO, SIMO, and MISO systems as special cases of the MIMO system. In the following sections, we focus on the MIMO systems with n_t transmit

antennas and n_r receive antennas. The channel model for these MIMO systems is illustrated in Fig. 2–1.

2.1.2 MIMO Channel Model

In this section, we discuss frequency flat fading and frequency selective fading MIMO channels. The corresponding input-output relations will be discussed. We consider a linear, discrete-time MIMO system with n_t transmit antennas and n_r receive antennas. Usually MIMO systems can be easily expressed by matrix-algebraic framework. In the following sections, matrix formulations of MIMO systems for both fading channel models will be introduced.

2.1.2.1 Frequency flat fading MIMO channel

A channel is classified as frequency flat fading, also known as frequency non-selective fading, if the bandwidth of the transmitted signal is much less than the coherence bandwidth of the channel. This implies that all the frequency components of the transmitted signal would roughly undergo the same degree of attenuation and phase shift. We can assume that only one copy of the signal is received. Slow fading also assumes that the channel coefficients are constant during the transmission of large number of symbols [38].

Let $h_{i,j}$ be a complex number corresponding to the channel coefficient from transmit antenna i to receive antenna j. If at a certain time instant signals $\{s_1, \ldots, s_{n_t}\}$ are transmitted from the n_t transmit antennas, the received signal at receive antenna

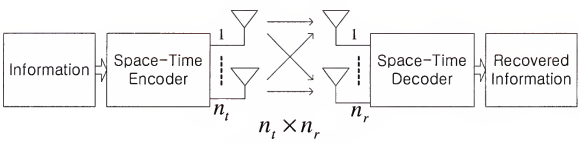


Figure 2–1: MIMO channel consisting of n_t transmit n_r receive antennas.

j can be expressed [24] as

$$y_j = \sum_{i=1}^{n_t} h_{i,j} s_i + e_j, \tag{2.1}$$

where e_j is the thermal noise at receive antenna j and e_j is assumed to be a zero-mean, circular-symmetric, complex Gaussian random variable with variance σ_e^2.

Let $\mathbf{s}$ and $\mathbf{y}$ be the n_t and n_r vectors containing the transmitted and received signals, respectively. Define the $n_r \times n_t$ channel matrix $\mathbf{H}$, which can be rewritten as

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix} \tag{2.2}$$

Thus (2.1) can be expressed by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{e}, \tag{2.3}$$

where $\mathbf{e} = [e_1, \ldots, e_{n_r}]^T$ is the noise sample vector at the receive antennas. Thus received signals during the time interval N can be easily expressed in a matrix form by

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{E}, \tag{2.4}$$

where the $n_r \times N$ matrix $\mathbf{Y} = [\mathbf{y_1}, \ldots, \mathbf{y}_N]$, $n_t \times N$ matrix $\mathbf{S} = [\mathbf{s_1}, \ldots, \mathbf{s}_N]$, and $n_r \times N$ matrix $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_N]$.

2.1.2.2 Frequency selective fading MIMO channel

A channel is classified frequency-selective if the bandwidth of the transmitted signal is large compared with the coherence bandwidth. In this case, different frequency components of the signal would undergo different degrees of fading. As the delays between different paths can be relatively large with respect to the symbol

duration, we would receive multiple copies of the signal [38]. Under this frequency-selective fading channel assumption, a MIMO system channel can be modeled as a causal-valued FIR filter [24] by

$$\mathbf{H}(z^{-1}) = \sum_{l=0}^{L} \mathbf{H}_l z^{-l}, \tag{2.5}$$

where L is the delay-spread of the channel and $\mathbf{H}_l$ is the $n_r \times n_t$ MIMO channel matrx for $l = 0, \ldots, L$. The transfer function associated with $\mathbf{H}(z^{-1})$ is given as

$$\mathbf{H}(\omega) = \sum_{l=0}^{L} \mathbf{H}_l e^{-j\omega l}. \tag{2.6}$$

Since the channel has memory, sequentially transmitted symbols will interfere with each other at the receiver; thus we need to consider a block, or sequence of symbols. In block transmission, we assume that transmitted sequence is preceded by a preamble and followed by a postamble. The preamble and postamble can be used for channel estimation and for preventing subsequent blocks from interfering with each other. If several blocks follow each other, the postamble of one block may also function as the preamble of the following block. It is also possible that blocks are separated with a guard interval. These two cases are illustrated in Fig. 2–2.

Consider the transmission of a given block, which is illustrated in Fig. 2–3. Let N_0, N_{pre}, and N_{post} be the length of the data, length of the preamble, and length of the postamble, respectively. To prevent the data of a preceding burst from interfering with the data part of the block under consideration, we must have $N_{pre} \geq L$. The received signal is well defined for $n = L - N_{pre}, \ldots, N_0 + N_{post} - 1$, but it depends on the transmitted data only for $n = 0, \ldots, N_0 + L - 1$.

Therefore the received data is given [24] by

$$\mathbf{y}(n) = \sum_{l=0}^{L} \mathbf{H}_l \mathbf{s}(n-l) + \mathbf{e}(n), \tag{2.7}$$

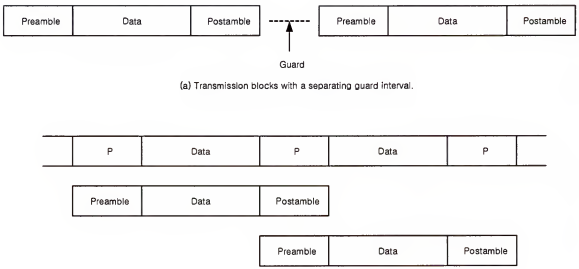


Figure 2–2: Block transmission [24].

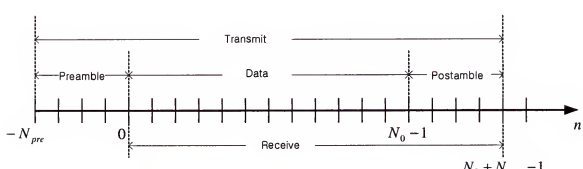


Figure 2–3: Definition of time index for the block transmission [24].

for $n = 0, \ldots, N_0 + L - 1$.

Let

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_L & \mathbf{H}_{L-1} & \cdots & \mathbf{H}_1 & \mathbf{H}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{H}_L & \ddots & & & \mathbf{H}_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{H}_L & \mathbf{H}_{L-1} & \cdots & \mathbf{H}_1 & \mathbf{H}_0 \end{bmatrix} \tag{2.8}$$

and

$$\begin{aligned} \mathbf{s} &= [\mathbf{s}^T(-L) \cdots \mathbf{s}^T(N_0 + L - 1)]^T \\ \mathbf{y} &= [\mathbf{y}^T(0) \cdots \mathbf{y}^T(N_0 + L - 1)]^T \\ \mathbf{e} &= [\mathbf{e}^T(0) \cdots \mathbf{e}^T(N_0 + L - 1)]^T. \end{aligned} \tag{2.9}$$

By using (2.8) and (2.9), we can express (2.7) by

$$\mathbf{y} = \bar{\mathbf{H}}\mathbf{s} + \mathbf{e}. \tag{2.10}$$

2.2 Minimum Variance Unbiased Estimation

In this section, we discuss estimators for an estimation of unknown deterministic parameters. We will focus our attention on the estimators which on the average yield the true parameter value. This class of estimators is known as unbiased estimators. We find unbiased estimators which can yield estimated values close to the true values with minimum variance [25]. Most of the information in this section is summarized from Kay [25].

2.2.1 Unbiased Estimator

An estimator is defined as an unbiased estimator if the estimator can yield the true value of the unknown parameter on the average. In other words, if the expected value of the estimator is the parameter being estimated, the estimator is said to be

unbiased. This can be expressed as

$$E(\hat{\theta}) = \theta, \tag{2.11}$$

where $\hat{\theta}$ is the estimated value and θ is the true value.

2.2.2 Minimum Variance Criterion

The quality of the estimator is usually measured by computing the mean square error (MSE) and it is defined as

$$\mathrm{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]. \tag{2.12}$$

This measures the average mean squared deviation of the estimator from the true value. In particular, if the estimator is unbiased, then the MSE of $\hat{\theta}$ is simply the variance of θ. Unfortunately this minimum MSE generally leads to unrealizable estimators because the MSE is composed of errors due to the variance of the estimators as well as the bias. Thus we need to constrain the bias to be zero and find the estimator which minimizes the variance. Such an estimator is known as the minimum variance unbiased (MVU) estimator [25].

2.2.3 Linear Models

As we discussed earlier, the minimum MSE approach generally leads to unrealizable estimators. MVU estimators does not exist in general. Thus we usually restrict the estimator to be linear in the data. In general, we can easily find a linear estimator that is unbiased and has the minimum variance.

Theorem 1. *[25] (MVU Estimator for the Linear Model with White Gaussian Noise) If the observed data is expressed as*

$$\mathbf{y} = \mathbf{S}\boldsymbol{\theta} + \mathbf{n}, \tag{2.13}$$

where $\mathbf{y}$ *is an* $N \times 1$ *vector,* $\mathbf{S}$ *is a known* $N \times p$ *matrix with* $N > p$ *and rank* p, $\boldsymbol{\theta}$ *is a* $p \times 1$ *vector of parameters to be estimated, and* $\mathbf{n}$ *is an* $N \times 1$ *white Gaussian*

noise vector with $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I})$. *Then the MVU estimator is*

$$\hat{\boldsymbol{\theta}} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{y} \tag{2.14}$$

where $(\cdot)^H$ *is complex conjugate transpose (Hermitian) of a matrix. The covariance matrix of* $\hat{\boldsymbol{\theta}}$ *is given as*

$$\mathbf{C}_{\hat{\theta}} = \sigma_n^2 (\mathbf{S}^H \mathbf{S})^{-1}. \tag{2.15}$$

Proof. See Chapter 4. of Kay [25]. □

We can easily verify that $\hat{\boldsymbol{\theta}}$ is unbiased by

$$\begin{aligned} E[\hat{\boldsymbol{\theta}}] &= E[(\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{y}] \\ &= E[(\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H (\mathbf{S}\boldsymbol{\theta} + \mathbf{n})] \\ &= E[\boldsymbol{\theta} + (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{n}] \\ &= \boldsymbol{\theta}. \end{aligned} \tag{2.16}$$

Hence

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma_n^2 (\mathbf{S}^H \mathbf{S})^{-1}). \tag{2.17}$$

Theorem 2. *[25] (MVU Estimator for the Linear Model with Colored Gaussian Noise)*

If the observed data is expressed as

$$\mathbf{y} = \mathbf{S}\boldsymbol{\theta} + \mathbf{n} \tag{2.18}$$

and the noise is distributed as

$$\mathbf{n} \sim \mathcal{N}(0, \mathbf{Q}), \tag{2.19}$$

where $\mathbf{Q}$ *is a positive definite noise covariance matrix. The MVU estimator is given as*

$$\hat{\boldsymbol{\theta}} = (\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{y} \tag{2.20}$$

and the covariance of $\hat{\boldsymbol{\theta}}$ *is given as*

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{S})^{-1} \tag{2.21}$$

Proof. See Chapter 4. of Kay [25]. □

We can easily verify that $\hat{\boldsymbol{\theta}}$ is unbiased by

$$\begin{aligned} E[\hat{\boldsymbol{\theta}}] &= E[(\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{y}] \\ &= E[(\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^H\mathbf{Q}^{-1}(\mathbf{S}\boldsymbol{\theta}+\mathbf{n})] \\ &= E[\boldsymbol{\theta} + (\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{n}] \\ &= \boldsymbol{\theta} \end{aligned} \tag{2.22}$$

Hence

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, (\mathbf{S}^H\mathbf{Q}^{-1}\mathbf{S})^{-1}) \tag{2.23}$$

2.2.4 Best Linear Unbiased Estimator

In practice, the MVU estimator often cannot be found even if it exists. As a result, suboptimal estimators can be used instead of the optimal MVU estimator if they meet the system requirements. Adopting linear model can be the solution to this problem. The best linear unbiased estimator (BLUE), which can be determined with knowledge of only the first and the second moments of the PDF, is one such estimator that we can consider. In general, the BLUE is more suitable for practical implementation [25].

Let the observed data set $\mathbf{y} = [y[0], y[1], \ldots, y[N-1]]^T$ and $\boldsymbol{\theta}$ be a $P \times 1$ vector which need to be estimated. Then an estimator that is linear in data is given as

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} y[n], \tag{2.24}$$

for $i = 1, 2, \ldots, p$ and a_{in}'s are weighting coefficients to be determined. a_{in} can be properly chosen to yield the BLUE. The BLUE will be optimal only when the MVU estimator turns out to be linear. To determine the BLUE, we have to find an estimator that is linear and unbiased and then determine the value of a_{in}'s which can minimize the variance of the estimator. The matrix form for (2.24) is given as

$$\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{y}, \tag{2.25}$$

where $\mathbf{A}$ is a $p \times N$ matrix. Same as the scalar parameter case, the unbiased constraint required is

$$E[\hat{\boldsymbol{\theta}}] = \mathbf{A}E[\mathbf{y}] = \boldsymbol{\theta}. \tag{2.26}$$

Thus we must have

$$E[\mathbf{y}] = \mathbf{S}\boldsymbol{\theta}, \tag{2.27}$$

where $\mathbf{S}$ is a known $N \times p$ matrix. From (2.26) and (2.27), we have the unbiased constraint which is given as

$$\mathbf{AS} = \mathbf{I}. \tag{2.28}$$

Let $\mathbf{a}_i = [a_{i0}, a_{i1}, \ldots, a_{i(N-1)}]$ for $i = 1, 2, \ldots, p$, then

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_p^T \end{bmatrix} \tag{2.29}$$

and let $\mathbf{s}_i$ be the ith column of $\mathbf{S}$, we have

$$\mathbf{S} = [\ \mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_p\ ]. \tag{2.30}$$

By using (2.29) and (2.30), the unbiased constraint of (2.28) reduces to

$$\mathbf{a}_i^T \mathbf{s}_j = \delta_{ij}, \tag{2.31}$$

for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$. Therefore the minimization problem to find BLUE is reduced to following optimization problem:

$$\begin{aligned} \min_{\mathbf{a}_i} \quad & var(\hat{\theta}_i) = \mathbf{a}_i^T \mathbf{Q} \mathbf{a}_i \\ \text{subject to} \quad & \mathbf{a}_i^T \mathbf{s}_j = \delta_{ij}, \end{aligned} \tag{2.32}$$

for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, p$. The solution is obtained and given in Kay [25] as

$$\mathbf{a}_{i_{opt}} = \mathbf{Q}^{-1}\mathbf{S}(\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{e}_i, \tag{2.33}$$

where $\mathbf{e}_i$ denotes the vector of all zeros except in the ith place. From this solution, we can obtain the BLUE for the vector parameter as

$$\hat{\boldsymbol{\theta}} = (\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{y} \tag{2.34}$$

and the covariance matrix of the estimator as

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{S})^{-1}. \tag{2.35}$$

We note from (2.34) and (2.35) that BLUE is the identical to the MVU estimator for the general linear model case in (2.20) and (2.21). As a result, we can conclude that for Gaussian data with the general linear model forms the BLUE is also the MVU

estimator with minimum variance

$$var(\hat{\theta}_i) = [(\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{S})^{-1}]_{ii}, \tag{2.36}$$

for $i = 1, 2, \ldots, p$.

2.2.5 Maximum Likelihood Estimator

The maximum likelihood estimator (MLE) is the estimator which is based on the maximum likelihood principle. The MLE is the most popular estimator due to its good performance for large data record, i.e., asymptotic efficiency and its close performance to the MVU estimator. In general, the MLE has the asymptotic properties of being unbiased, achieving CRLB, and having a Gaussian error PDF [25].

The basic idea to find the MLE is finding the value of θ that maximizes the likelihood function $p(\mathbf{y};\theta)$ for fixed $\mathbf{y}$. In this section, we focus only on the general linear data model in the vector parameter case. Let us consider the linear data model given as

$$\mathbf{y} = \mathbf{S}\boldsymbol{\theta} + \mathbf{n}, \tag{2.37}$$

where $\mathbf{y}$ is an $N \times 1$ observed data vector, $\mathbf{S}$ is a known $N \times p$ matrix with $N > p$, $\boldsymbol{\theta}$ is a $p \times 1$ parameter vector to be estimated, and $\mathbf{n}$ is an $N \times 1$ noise vector with PDF $\mathcal{N}(0, \mathbf{Q})$ given by

$$p(\mathbf{y};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}}\det^{\frac{1}{2}}(\mathbf{Q})} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{S}\boldsymbol{\theta})^T\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{S}\boldsymbol{\theta})\right]. \tag{2.38}$$

Therefore the MLE of the $\boldsymbol{\theta}$ can be found by maximizing the PDF. This is the same as minimizing $(\mathbf{y} - \mathbf{S}\boldsymbol{\theta})^T\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{S}\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Then the MLE of $\boldsymbol{\theta}$ is given [25] as

$$\hat{\boldsymbol{\theta}} = (\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{y} \tag{2.39}$$

and the covariance matrix of the estimator is given as

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{S}^T\mathbf{Q}^{-1}\mathbf{S})^{-1}. \tag{2.40}$$

We note that this $\hat{\boldsymbol{\theta}}$ is efficient and is the MVU estimator. Thus the MLE is identical to the BLUE and is an optimal estimator in general linear data model with Gaussian noise.

2.3 Bayesian Estimator

In the classical approach to statistical estimation, we assume the parameter θ that need to be estimated is a deterministic but unknown constant. Contrary to the classical approach, in the Bayesian approach, we assume that θ is a random variable with a given prior PDF. Thus the Bayesian approach can improve the estimation accuracy by using prior knowledge of θ. Moreover Bayesian estimation is useful in situations where the MVU estimator cannot be found [25]. Our goal in this section is to find an estimator $\hat{\theta}$ that can minimize the Bayesian MSE. To do so, we need to define the Bayesian MSE as

$$\text{Bmse}(\hat{\theta}) = E[(\theta - \hat{\theta})^2]. \tag{2.41}$$

Note that the expectation in (2.41) is with respect to the joint PDF $\text{p}(\mathbf{y}, \theta)$. It is easily seen that the optimal estimator in terms of minimizing the Bayesian MSE is the mean of the poterior PDF $\text{p}(\theta|\mathbf{y})$ or $\hat{\theta}$=E$(\theta|\mathbf{y})$ and the Bayesian MSE is just the variance of the posterior PDF when averaged over the PDF of $\mathbf{y}$ [25]. The term posterior PDF refers to the PDF of θ after the data have been observed. In contrast, the prior PDF indicates the PDF before the data are observed. We also call the estimator that minimizes the Bayesian MSE, the minimum mean square error (MMSE) estimator. Let the observed data be modeled as

$$\mathbf{y} = \mathbf{S}\theta + \mathbf{n}, \tag{2.42}$$

where $\mathbf{y}$ is an N $\times$ 1 data vector, $\mathbf{S}$ is a known N$\times$ p matrix, θ is a p$\times$ 1 random vector with prior PDF $\mathcal{N}(\mu_\theta,\mathbf{C}_\theta)$, and $\mathbf{n}$ is an N$\times$ 1 noise vector with PDF $\mathcal{N}(0,\mathbf{C}_n)$ and independent of θ. This Bayesian general linear model is different from the classical general linear model in that θ is modeled as a random variable with a Gaussian prior PDF. With the system model in (2.42) and the assumptions we made, the posterior PDF p($\theta|\mathbf{y}$)is Gaussian with mean [25]

$$E(\theta|\mathbf{y}) = \mu_\theta + \mathbf{C}_\theta\mathbf{S}^T(\mathbf{S}\mathbf{C}_\theta\mathbf{S}^T + \mathbf{C}_n)^{-1}(\mathbf{y} - \mathbf{S}\mu_\theta) \tag{2.43}$$

and covariance

$$\mathbf{C}_{\theta|y} = \theta - \mathbf{C}_\theta\mathbf{S}^T(\mathbf{S}\mathbf{C}_\theta\mathbf{S}^T + \mathbf{C}_n)^{-1}\mathbf{S}\mathbf{C}_\theta. \tag{2.44}$$

By applying the matrix inversion lemma, (2.43) and (2.44) can be expressed in alternate form as

$$E(\theta|\mathbf{y}) = \mu_\theta + (\mathbf{C}_\theta^{-1} + \mathbf{S}^T\mathbf{C}_n^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{C}_n^{-1}(\mathbf{y} - \mathbf{S}\mu_\theta) \tag{2.45}$$

and

$$\mathbf{C}_{\theta|y} = (\mathbf{C}_\theta^{-1} + \mathbf{S}^T\mathbf{C}_n^{-1}\mathbf{S})^{-1}. \tag{2.46}$$

Note that contrary to the classical general linear model, the known matrix $\mathbf{S}$ need not be of full rank to insure the invertibility $(\mathbf{S}\mathbf{C}_\theta\mathbf{S}^T + \mathbf{C}_n)$ in (2.43) and (2.44). One interesting fact is that if there is no prior knowledge, the Bayesian estimator yields the same form as the MVU estimator for the classical linear model. It can be easily verified from (2.45) with no prior knowledge $\mathbf{C}_\theta^{-1} = 0$, and therefore

$$\hat{\theta} = (\mathbf{S}^T\mathbf{C}_n^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{C}_n^{-1}\mathbf{y}, \tag{2.47}$$

which is the same as the MVU estimator for the general linear model. What is the meaning of the condition $\mathbf{C}_\theta^{-1} = 0$ in above? If we assume the elements of θ are

uncorrelated, then $\mathbf{C}_\theta$ is a diagonal matrix, with diagonal elements $\sigma^2_{\theta,i}$. When all of these variances are very large, $\mathbf{C}_\theta^{-1} \approx 0$. A large variance for θ_i means we have no idea where θ_i is located about its mean value [26]. The following theorem summarizes the results of this section.

Theorem 3. *[25] If we consider the Bayesian linear model in (2.42), the MMSE estimator is,*

$$\hat{\theta} = \mu_\theta + (\mathbf{C}_\theta^{-1} + \mathbf{S}^T\mathbf{C}_n^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{C}_n^{-1}(\mathbf{y} - \mathbf{S}\mu_\theta) \tag{2.48}$$

and the performance of the estimator is measured by the error $\epsilon = \theta - \hat{\theta}$, whose PDF is Gaussian with mean zero and covariance matrix

$$\mathbf{C}_\epsilon = (\mathbf{C}_\theta^{-1} + \mathbf{S}^T\mathbf{C}_n^{-1}\mathbf{S})^{-1}. \tag{2.49}$$

Proof. See Chapter 8. of Kay [25]. □

2.4 Circulant Matrices and Toeplitz Matrices

In this section, we briefly discuss the properties of circulant matrices and Toeplitz matrices. When a random processes is wide-sense stationary, its covariance matrix has the form of a Toeplitz matrix. Because circulant matrices are used to approximate and explain the behavior of the Toeplitz matrices, it is helpful to review the properties and the relation between two types of matrices. We also discuss the asymptotic behavior of these two types of matrices. Information in this section is summarized from Gray [27].

2.4.1 Circulant Matrices

A circulant matrix $\mathbf{C}$ has the form given as

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ \vdots & c_{n-1} & c_0 & \ddots & \vdots \\ c_2 & \cdots & \ddots & \ddots & c_1 \\ c_1 & c_2 & \cdots & c_{n-1} & c_0 \end{bmatrix} \tag{2.50}$$

where each row is a cyclic shift of the row above it.

The eigenvalues λ_m and the eigenvectors $v^{(m)}$ of $\mathbf{C}$ are the solutions of

$$\mathbf{C}v = \lambda v, \tag{2.51}$$

where the eigenvalues and eigenvectors are given by

$$\lambda_m = \sum_{k=0}^{n-1} c_k e^{-j\frac{2\pi mk}{n}} \tag{2.52}$$

and

$$v^{(m)} = \frac{1}{\sqrt{n}}(1, e^{-j\frac{2\pi m}{n}}, \cdots, e^{-j\frac{2\pi(n-1)}{n}}), \tag{2.53}$$

for $m = 0, 1, \ldots, n-1$. From (2.52) and (2.53), we can write

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \tag{2.54}$$

where

$$[\mathbf{U}]_{mk} = \frac{1}{\sqrt{n}} e^{-j\frac{2\pi mk}{n}}, \quad \text{for } m, k = 0, 1, \ldots, n-1$$

and $\mathbf{\Lambda}$ is a diagonal matrix with elements $\lambda_k \delta_{k-j}$ where δ is the Kronecker delta function defined as

$$\delta_m = \begin{cases} 1 & m = 0 \\ 0 & otherwise \end{cases}$$

Note that any matrix that can be expressed in the form of (2.54) is a circulant matrix. In addition, λ_m in (2.52) is simply the discrete Fourier transform of the sequence c_k. Note that (2.54) can be interpreted as a combination of the inverse Fourier formula and the Fourier cyclic shift formula. Moreover, all circulant matrices have the same set of eigenvectors [27].

2.4.2 Toeplitz Matrices

An $n \times n$ matrix $\mathbf{T}_n$ is Toeplitz matrix with elements $t_{k,j}$ where $t_{k,j} = t_{k-j}$ and has the form

$$\mathbf{T}_n = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & \cdots & t_{-(n-2)} \\ \vdots & t_1 & t_0 & \ddots & \vdots \\ t_{n-2} & \cdots & \ddots & \ddots & t_{-1} \\ t_{n-1} & t_{n-2} & \cdots & t_1 & t_0 \end{bmatrix} \tag{2.55}$$

Example of such matrices are covariance matrices of wide-sense stationary random process and matrix representations of linear time-invariant discrete time filters [27]. Consider the infinite sequence t_k for $k = 0, \pm 1, \pm 2, \cdots$ and define the finite $n \times n$ Toeplitz matrix $\mathbf{T}_n$ as in (2.55). Toeplitz matrices can be categorized by the restrictions placed on the sequence t_k. $\mathbf{T}_n$ is said to be a finite order Toeplitz matrix if there exists a finite m such that $t_k = 0$, $|k| > m$. If t_k is an infinite sequence, then there are two common constraints. The most general assumption is that the t_k are square summable that is

$$\sum_{k=-\infty}^{\infty} |t_k|^2 < \infty \tag{2.56}$$

and the other stronger assumption that t_k are absolutely summable that is

$$\sum_{k=-\infty}^{\infty} |t_k| < \infty. \tag{2.57}$$

Since

$$\sum_{k=-\infty}^{\infty} |t_k|^2 \leq \left\{ \sum_{k=-\infty}^{\infty} |t_k| \right\}^2, \tag{2.58}$$

we note that (2.57) is indeed stronger constraint than (2.56). We only consider the stronger assumption case, i.e., t_k are absolutely summable, because it can simplify the mathematics but does not alter the fundamental concepts involved [27]. Another main advantage of (2.57) over (2.56) is that it ensures the existence and continuity of the Fourier series $f(\tau)$ which is defined by

$$\begin{aligned} f(\tau) &= \sum_{k=-\infty}^{\infty} t_k e^{jk\tau} \\ &= \lim_{n\to\infty} \sum_{k=-n}^{n} t_k e^{jk\tau}. \end{aligned} \tag{2.59}$$

2.4.3 Absolutely Summable Toeplitz Matrix

If we have absolutely summable sequence t_k with

$$f(\tau) = \sum_{k=-\infty}^{\infty} t_k e^{jk\tau}, \tag{2.60}$$

t_k is obtained as

$$t_k = \frac{1}{2\pi} \int_0^{2\pi} f(\tau) e^{-jk\lambda}, \tag{2.61}$$

for $k = 0, \pm 1, \pm 2, \cdots$.

Define $\mathbf{C}_n(f)$ to be the circulant matrix with top row $(c_0^{(n)}, c_1^{(n)}, \ldots, c_{n-1}^{(n)})$ where

$$c_k^{(n)} = n^{-1} \sum_{i=0}^{n-1} f\left(\frac{2\pi i}{n}\right) e^{j\frac{2\pi ik}{n}}. \tag{2.62}$$

For fixed k, we have

$$\begin{aligned}\lim_{n\to\infty} c_k^{(n)} &= \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} f\left(\frac{2\pi i}{n}\right) e^{j\frac{2\pi ik}{n}} \\ &= (2\pi)^{-1} \int_0^{2\pi} f(\tau) e^{jk\tau} d\tau \\ &= t_{-k}. \end{aligned} \tag{2.63}$$

From (2.52) and (2.63), the eigenvalues of $\mathbf{C}_n$ are simply $f(\frac{2\pi m}{n})$ [27] and they are given by

$$\begin{aligned}\lambda_{n,m} &= \sum_{k=0}^{n-1} c_k^{(n)} e^{-j\frac{2\pi mk}{n}} \\ &= \sum_{k=0}^{n-1} \left(n^{-1} \sum_{i=0}^{n-1} f\left(\frac{2\pi i}{n}\right) e^{j\frac{2\pi ik}{n}} \right) e^{-j\frac{2\pi mk}{n}} \\ &= \sum_{i=0}^{n-1} f\left(\frac{2\pi i}{n}\right) \left(n^{-1} \sum_{k=0}^{n-1} e^{j\frac{2\pi k(i-m)}{n}} \right) \\ &= f\left(\frac{2\pi m}{n}\right). \end{aligned} \tag{2.64}$$

If $\mathbf{C}_n$ is a circulant matrix with eigenvalues $f\left(\frac{2\pi m}{n}\right)$ for $m = 0, 1, \ldots, n-1$, then

$$\begin{aligned} c_k^{(n)} &= n^{-1} \sum_{m=0}^{n-1} \lambda_{n,m} e^{j\frac{2\pi mk}{n}} \\ &= n^{-1} \sum_{m=0}^{n-1} f\left(\frac{2\pi m}{n}\right) e^{j\frac{2\pi mk}{n}}, \end{aligned} \tag{2.65}$$

as in (2.62). We can use either (2.62) or (2.65) to define $\mathbf{C}_n(f)$.

Lemma 1 : [27] Given the function $f(\tau)$ in (2.61) and the circulant matrix $\mathbf{C}_n(f)$ defined by (2.62), then

$$c_k^{(n)} = \sum_{m=-\infty}^{\infty} t_{-k+mn}, \tag{2.66}$$

for $k = 0, 1, \ldots, n-1$.

(Note that the sum exists since the t_k are absolutely summable.)

■

From (2.66), we note the shortcoming of $\mathbf{C}_n(f)$ for applications as a circulant approximation to $\mathbf{T}_n(f)$ is that it depends on the entire sequence $\{t_k; k = 0, \pm 1, \pm 2, \cdots\}$ and not just on the finite elements $\{t_k; k = 0, \pm 1, \pm 2, \cdots \pm n-1\}$ of $\mathbf{T}_n(f)$. This can cause problems where we wish a circulant approximation to a Toeplitz matrix $\mathbf{T}_n$ when we only know $\mathbf{T}_n$ and not f.

An approximation is to form the truncated Fourier series [27]

$$\hat{f}_n(\tau) = \sum_{k=-n}^{n} t_k e^{jk\tau}, \tag{2.67}$$

which depends only on $\{t_k; k = 0, \pm 1, \pm 2, \cdots \pm n-1\}$, and define a circulant matrix as

$$\hat{\mathbf{C}}_n = \mathbf{C}_n(\hat{f}_n). \tag{2.68}$$

The circulant matrix is with top row $(\hat{c}_0^{(n)}, \ldots, \hat{c}_{n-1}^{(n)})$, where

$$\begin{aligned} \hat{c}_k^{(n)} &= n^{-1} \sum_{i=0}^{n-1} \hat{f}_n \left(\frac{2\pi i}{n} \right) e^{j\frac{2\pi ik}{n}} \\ &= n^{-1} \sum_{i=0}^{n-1} \left(\sum_{m=-n}^{n} t_m e^{\frac{j2\pi im}{n}} \right) e^{j\frac{2\pi ik}{n}} \\ &= \sum_{m=-n}^{n} t_m \left(n^{-1} \sum_{i=0}^{n-1} e^{j\frac{2\pi i(k+m)}{n}} \right). \end{aligned} \tag{2.69}$$

If $m = -k$ or $m = n - k$ in the last term of (2.69), $\hat{c}_k^{(n)}$ is given as

$$\hat{c}_k^{(n)} = t_{-k} + t_{n-k}, \tag{2.70}$$

for $k = 0, 1, \ldots, n-1$.

This result will be applied to our feedback design scheme. Finally, the following lemma shows that these circulant matrices are asymptotically equivalent to each other and to $\mathbf{T}_n$.

Lemma 2; [27] Let $\mathbf{T}_n$ with elements t_{k-j} where

$$\sum_{k=-\infty}^{\infty} |t_k| < \infty \tag{2.71}$$

and define

$$f(\tau) = \sum_{k=-\infty}^{\infty} t_k e^{jk\tau}. \tag{2.72}$$

Define the circulant matrices $\mathbf{C}_n(f)$ and $\hat{\mathbf{C}}_n = \mathbf{C}_n(\hat{f}_n)$ as in (2.62), (2.67), and (2.68). Then,

$$\mathbf{C}_n(f) \sim \hat{\mathbf{C}}_n \sim \mathbf{T}_n. \tag{2.73}$$

■

Proof. See Chapter 4. of Gray [27]. □

2.5 Summary

In this chapter, we introduced MIMO wireless communication systems with antenna arrays at both the transmitter and receiver. We briefly introduced channel and signal models for the MIMO systems. Moreover, we summarized the approach of minimum variance unbiased estimation. The BLUE and ML channel estimator were discussed. In addition, we introduced Bayesian estimator and its properties. Finally, we discussed the properties of circulant matrices and Toeplitz matrices. In particular, we studied asymptotic behavior of these two types of matrices.

CHAPTER 3
SYSTEM MODEL AND CHANNEL ESTIMATION

In this chapter, we describe the MIMO system model. We also develop the best linear unbiased estimator (BLUE) and Bayesian estimator for the MIMO channels. In addition, we obtain the mean square errors (MSEs) of the BLUE and Bayesian channel estimator. In particular, in the BLUE approach, we consider both the cases where the interference covariance matrix is non-singular and singular. The expressions of the BLUE for both conditions are derived.

3.1 Channel Estimation with BLUE

3.1.1 System Model

We consider a single-user MIMO system with n_t transmit antennas and n_r receive antennas over a frequency flat fading channel in the presence of colored interference. We assume that the colored interference is composed of a jamming signal transmitted by n_j transmit antennas. We assume that the transmission from the transmitter to the receiver is packetized. Each packet contains a training frame that is composed of a set of known training sequences, each of which is sent out by a transmit antenna. The observed training symbols at the receiver for a packet are given by

$$\mathbf{Y} = \mathbf{HS} + \mathbf{H}_J\mathbf{S}_J, \tag{3.1}$$

where $\mathbf{S}$ is the $n_t \times N$ transmitted training symbol matrix that is known to the receiver, N is the number of training symbols, and $\mathbf{S}_J$ is the $n_j \times N$ jamming signal matrix. We assume that symbols in $\mathbf{S}_J$ are identically distributed zero-mean random variables, independent across space and correlated across time. We assume that the number of training symbols N is larger than n_t. The $n_r \times n_t$ matrix $\mathbf{H}$ and $n_r \times n_j$ matrix $\mathbf{H}_J$ are channel matrices from the transmitter and the jammer to the receiver,

respectively. We assume that both $\mathbf{H}$ and $\mathbf{H}_J$ are unknown but deterministic for the channel estimation problem considered in this paper. Moreover, we assume that the power of the thermal noise is much smaller than that of the signal and jammer, and hence the effect of the thermal noise is ignored in the above formulation [28].

Let $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{h} = \text{vec}(\mathbf{H})$, and $\mathbf{e} = \text{vec}(\mathbf{H}_J\mathbf{S}_J)$. Taking transpose and then vectorizing on both sides of (3.1), we have [29]

$$\mathbf{y} = (\mathbf{S}^T \otimes \mathbf{I}_{n_r})\mathbf{h} + \mathbf{e}. \tag{3.2}$$

We note that if $\mathbf{S}$ is not of full (column) rank, the projection of $\mathbf{h}$ in the null space of $\mathbf{S}^T \otimes \mathbf{I}_{n_r}$ will not be observable[1] from $\mathbf{y}$. Hence, we impose the restriction that $\mathbf{S}$ is of full rank.

From the channel model described in (3.1), we note that the correlation matrix of the jammer vector $\mathbf{Q} = E[\mathbf{e}\mathbf{e}^H]$ decomposes into a Kronecker product $\mathbf{Q} = \mathbf{Q}_N \otimes \mathbf{Q}_r$, where $\mathbf{Q}_N$ is an $N \times N$ matrix and $\mathbf{Q}_r$ is an $n_r \times n_r$ matrix, representing the correlations of the noise in time and space, respectively. To see this, write the $n_j \times N$ jamming signal matrix as $\mathbf{S}_J = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N]$, where $\mathbf{s}_i$ is the $n_j \times 1$ vector transmitted by the jammer at time i. Since the elements of $\mathbf{S}_J$ are independent across space (rows), we have

$$E[\mathbf{s}_i\mathbf{s}_k^H] = R_J(i,k) \cdot \mathbf{I}_{n_j}, \tag{3.3}$$

where $R_J(i,k)$ is the time correlation between the ith and kth jamming symbols. Let $R_J(i,k)$ be the (i,k)th element of $\mathbf{Q}_N$, then

$$E[\text{vec}(\mathbf{S}_J)\text{vec}(\mathbf{S}_J)^H] = \mathbf{Q}_N \otimes \mathbf{I}_{n_j}. \tag{3.4}$$

---

[1] This notion is made more precise in Section 3.1.2.

Since $\mathbf{e} = \text{vec}(\mathbf{H}_J\mathbf{S}_J) = (\mathbf{I}_N \otimes \mathbf{H}_J)\text{vec}(\mathbf{S}_J)$ [29], we have

$$\begin{aligned}\mathbf{Q} &= (\mathbf{I}_N \otimes \mathbf{H}_J)E[\text{vec}(\mathbf{S}_J)\text{vec}(\mathbf{S}_J)^H](\mathbf{I}_N \otimes \mathbf{H}_J)^H \\ &= (\mathbf{I}_N \otimes \mathbf{H}_J)(\mathbf{Q}_N \otimes \mathbf{I}_{n_j})(\mathbf{I}_N \otimes \mathbf{H}_J)^H \\ &= \mathbf{Q}_N \otimes \underbrace{\mathbf{H}_J\mathbf{H}_J^H}_{\mathbf{Q}_r}, \end{aligned} \tag{3.5}$$

where the third equality is obtained by using $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ and $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$ [29]. We note that $\mathbf{Q}_N$ is non-singular under most practical scenarios. However, $\mathbf{Q}$ is not necessarily non-singular. For instance, when $n_j < n_r$, $\mathbf{Q}_r$ is singular, and hence $\mathbf{Q}$ is also singular.

We assume that $\mathbf{H}$ and $\mathbf{H}_J$ remain unchanged during the observation interval. In addition, we assume that $\mathbf{Q}_N$ varies at a rate that is much slower than that of $\mathbf{H}$ and $\mathbf{H}_J$. As a result, it is possible for the receiver to feed back information of $\mathbf{Q}_N$ to the transmitter so that it can make use of this information for the estimation of $\mathbf{H}$.

3.1.2 Best Linear Unbiased Channel Estimator

In this section, we develop the best linear unbiased estimator (BLUE) for the channel vector $\mathbf{h}$. Let us denote this BLUE by $\hat{\mathbf{h}}$. Then $\hat{\mathbf{h}}$ is optimal in the sense that it has the minimum total variance among all linear unbiased estimators for $\mathbf{h}$. That is the mean square estimation error defined by $\text{MSE} = E[\|\hat{\mathbf{h}} - \mathbf{h}\|^2]$ is minimized by the BLUE $\hat{\mathbf{h}}$.

3.1.2.1 Non-singular jammer covariance matrix Q

When the jammer covariance matrix $\mathbf{Q}$ is non-singular, it is introduced in Section 2.2.4 that the BLUE for $\mathbf{h}$, assuming $\mathbf{S}$ is of full rank, is given by,

$$\begin{aligned}\hat{\mathbf{h}} &= [(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H\mathbf{Q}^{-1}(\mathbf{S}^T \otimes \mathbf{I}_{n_r})]^{-1}(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H\mathbf{Q}^{-1}\mathbf{y} \\ &= [(\mathbf{S}^*\mathbf{Q}_N^{-1}\mathbf{S}^T)^{-1}\mathbf{S}^*\mathbf{Q}_N^{-1} \otimes \mathbf{I}_{n_r}]\mathbf{y}, \end{aligned} \tag{3.6}$$

where we have used the fact $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ [29] to obtain the second equality. We also note that if $\mathbf{e}$ is a Gaussian random vector, then $\hat{\mathbf{h}}$ given above is also the maximum likelihood estimator (MLE), introduced in Section 2.2.5 from Chapter 2, for $\mathbf{h}$. Writing (3.6) back into matrix form, we have

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{Q}_N^{-1}\mathbf{S}^H(\mathbf{S}\mathbf{Q}_N^{-1}\mathbf{S}^H)^{-1}. \tag{3.7}$$

Moreover, the MSE of the BLUE is given by

$$\begin{aligned}
\text{MSE} &= E[\|\hat{\mathbf{h}} - \mathbf{h}\|^2] \\
&= \text{tr}[(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H \mathbf{Q}^{-1} (\mathbf{S}^T \otimes \mathbf{I}_{n_r})]^{-1} \\
&= \text{tr}[(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H (\mathbf{Q}_N \otimes \mathbf{Q}_r)^{-1} (\mathbf{S}^T \otimes \mathbf{I}_{n_r})]^{-1} \\
&= \text{tr}[(\mathbf{S}^* \mathbf{Q}_N^{-1} \otimes \mathbf{I}_{n_r} \mathbf{Q}_r^{-1})(\mathbf{S}^T \otimes \mathbf{I}_{n_r})]^{-1} \\
&= \text{tr}[(\mathbf{S}^* \mathbf{Q}_N^{-1} \mathbf{S}^T) \otimes \mathbf{Q}_r^{-1}]^{-1} \\
&= \text{tr}[(\mathbf{S}^* \mathbf{Q}_N^{-1} \mathbf{S}^T)^{-1}] \text{tr}(\mathbf{Q}_r).
\end{aligned} \tag{3.8}$$

The second equality is obtained from $\|\mathbf{X}\|^2 = \text{tr}\{\mathbf{X}^H\mathbf{X}\}$. The fourth and fifth equalities are obtained by using $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ and $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$. The last equality is due to $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$ [29].

3.1.2.2 Singular jammer covariance matrix Q

In this section, we focus on the case of singular $\mathbf{Q}$. First, write the spectral decompositions of $\mathbf{Q}_N$ and $\mathbf{Q}_r$, respectively, in the following forms:

$$\begin{aligned}
\mathbf{Q}_N &= \underbrace{\begin{bmatrix} \tilde{\mathbf{U}}_N & \hat{\mathbf{U}}_N \end{bmatrix}}_{\mathbf{U}_N} \underbrace{\begin{bmatrix} \tilde{\mathbf{\Lambda}}_N & 0 \\ 0 & 0 \end{bmatrix}}_{\mathbf{\Lambda}_N} \begin{bmatrix} \tilde{\mathbf{U}}_N^H \\ \hat{\mathbf{U}}_N^H \end{bmatrix} = \tilde{\mathbf{U}}_N \tilde{\mathbf{\Lambda}}_N \tilde{\mathbf{U}}_N^H, \\
\mathbf{Q}_r &= \underbrace{\begin{bmatrix} \tilde{\mathbf{U}}_r & \hat{\mathbf{U}}_r \end{bmatrix}}_{\mathbf{U}_r} \underbrace{\begin{bmatrix} \tilde{\mathbf{\Lambda}}_r & 0 \\ 0 & 0 \end{bmatrix}}_{\mathbf{\Lambda}_r} \begin{bmatrix} \tilde{\mathbf{U}}_r^H \\ \hat{\mathbf{U}}_r^H \end{bmatrix} = \tilde{\mathbf{U}}_r \tilde{\mathbf{\Lambda}}_r \tilde{\mathbf{U}}_r^H,
\end{aligned} \tag{3.9}$$

where $\tilde{\mathbf{\Lambda}}_N$ and $\tilde{\mathbf{\Lambda}}_r$ are the diagonal matrices that contain the positive eigenvalues of $\mathbf{Q}_N$ and $\mathbf{Q}_r$, respectively. Then the spectral decomposition of $\mathbf{Q}$ is given [29] by

$$\begin{aligned}\mathbf{Q} &= (\mathbf{U}_N \otimes \mathbf{U}_r)(\mathbf{\Lambda}_N \otimes \mathbf{\Lambda}_r)(\mathbf{U}_N \otimes \mathbf{U}_r)^H \\ &= \begin{bmatrix}(\tilde{\mathbf{U}}_N \otimes \tilde{\mathbf{U}}_r) & \hat{\mathbf{U}}\end{bmatrix} \begin{bmatrix}\tilde{\mathbf{\Lambda}}_N \otimes \tilde{\mathbf{\Lambda}}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}(\tilde{\mathbf{U}}_N \otimes \tilde{\mathbf{U}}_r)^H \\ \hat{\mathbf{U}}^H\end{bmatrix}, \qquad (3.10)\end{aligned}$$

where the second equality is obtained by grouping the positive eigenvalues of $\mathbf{Q}$ together. Consider applying $\hat{\mathbf{U}}^H$ to the received vector $\mathbf{y}$:

$$\begin{aligned}\hat{\mathbf{U}}^H\mathbf{y} &= \hat{\mathbf{U}}^H(\mathbf{S}^T \otimes \mathbf{I}_{n_r})\mathbf{h} + \hat{\mathbf{U}}^H\mathbf{e} \\ &= \hat{\mathbf{U}}^H \underbrace{(\mathbf{S}^T \otimes \mathbf{I}_{n_r})}_{\mathbf{X}}\mathbf{h} \quad \text{with probability 1.} \qquad (3.11)\end{aligned}$$

Indeed, since $\mathcal{N}(\mathbf{Q}) = \mathcal{R}(\hat{\mathbf{U}})$, $\hat{\mathbf{U}}^H\mathbf{e}$ has both zero mean and zero covariance, and hence the second equality above results. The requirement in (3.11) imposes a constraint on the linear estimation problem. If we neglect those $\mathbf{y}$ (which occur with zero probability) that (3.11) is not satisfied, then a *consistent model* will result for the constrained estimation problem. We use the triple $(\mathbf{y}, \mathbf{Xh}, \mathbf{Q})$ to denote such a consistent model.

The constrained estimation problem mentioned above can be conveniently studied by employing the theory of generalized inverses [31]. First, we have to make sure a linear unbiased estimator for $\mathbf{h}$ exists under the constrained model $(\mathbf{y}, \mathbf{Xh}, \mathbf{Q})$. To do so, we need the following characterization [31, Ch. 6].

Definition 1. *A linear function* $\mathbf{c}^H\mathbf{h}$ *is referred to as* linearly unbiasedly estimable *under* $(\mathbf{y}, \mathbf{Xh}, \mathbf{Q})$ *if there exists a vector* $\mathbf{a}$ *and scalar* α *such that* $E[\mathbf{a}^H\mathbf{y} + \alpha] = \mathbf{c}^H\mathbf{h}$ *for all* $\mathbf{h}$ *such that* $\hat{\mathbf{U}}^H\mathbf{y} = \hat{\mathbf{U}}^H\mathbf{Xh}$.

With this characterization, the follow result specifies the set of all linear unbiased estimators of $\mathbf{h}$ under $(\mathbf{y}, \mathbf{Xh}, \mathbf{Q})$.

Theorem 4. *The function* $\mathbf{c}^H\mathbf{h}$ *is linearly unbiasedly estimable under* $(\mathbf{y}, \mathbf{Xh}, \mathbf{Q})$ *if and only if* $\mathbf{c} \in \mathcal{R}(\mathbf{X}^H)$. *Moreover, if this condition is satisfied, the set of linear unbiased estimators for* $\mathbf{c}^H\mathbf{h}$ *is given by* $\{\mathbf{a}^H\mathbf{h} : \mathbf{a}^H\mathbf{X} = \mathbf{c}^H\}$.

Proof. See the proofs of Theorems 6.4.1 and 6.4.2 in Campbell and Meyer [31]. □

Corollary 1. *Suppose that the training symbol matrix* $\mathbf{S}$ *is of full column rank. Then every* $\mathbf{Ay}$ *such that* $\mathbf{A}(\mathbf{S}^T \otimes \mathbf{I}_{n_r}) = \mathbf{I}_{n_t n_r}$ *is a linear unbiased estimator for the channel vector* $\mathbf{h}$.

Proof. Since $\mathbf{S}$ is of full column rank, $\mathbf{X} = \mathbf{S}^T \otimes \mathbf{I}_{n_r}$ is of full row rank [29]. Let $\mathbf{e}_i$, for $i = 1, 2, \ldots, n_t n_r$, denote the elementary vectors of dimension $n_t n_r$. Then they are all in $\mathcal{R}(\mathbf{X}^H)$. Thus, $\mathbf{e}_i^H\mathbf{h}$, for $i = 1, 2, \ldots, n_t n_r$, are linearly unbiasedly estimable, and the set of linear unbiased estimators for $\mathbf{h}$ is $\{\mathbf{Ay} : \mathbf{AX} = \mathbf{I}_{n_t n_r}\}$ by Theorem 4. □

Next, we turn to finding the BLUE for $\mathbf{h}$ with the assumption that $\mathbf{S}$ is of full column rank. To do so, we need to introduce some generalized inverses of a matrix [31].

Definition 2. *Moore-Penrose inverse and (1)-inverse:*

(a) The Moore-Penrose inverse, $\mathbf{A}^\dagger$, *of a matrix* $\mathbf{A}$ *is the unique matrix that satisfies:*

(i) $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$,

(ii) $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$,

(iii) $(\mathbf{A}\mathbf{A}^\dagger)^H = \mathbf{A}\mathbf{A}^\dagger$, *and*

(iv) $(\mathbf{A}^\dagger\mathbf{A})^H = \mathbf{A}^\dagger\mathbf{A}$.

(b) A matrix $\mathbf{A}^-$ *is called an (1)-inverse of* $\mathbf{A}$ *if* $\mathbf{A}^-$ *satisfies (i) above, i.e.,* $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$. *We denote the set of all (1)-inverses of* $\mathbf{A}$ *by* $\mathbf{A}\{1\}$. *Obviously,* $\mathbf{A}^\dagger \in \mathbf{A}\{1\}$.

Hence, from Corollary 1, we need to consider all estimators of the form $(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^{-}\mathbf{y}$. The following theorem provides a means to find the BLUE among these linear unbiased estimators.

Theorem 5. *Suppose that* $\mathbf{c}^H\mathbf{h}$ *is linearly unbiasedly estimable under* $(\mathbf{y}, \mathbf{Xh}, \mathbf{Q})$. *Let* $\mathbf{K}$, $\mathbf{L}$, *and* $\mathbf{M}$ *be three matrices that satisfy the following conditions:*

(i) $(\mathbf{I}_{Nn_r} - \mathbf{X}\mathbf{X}^{\dagger})\mathbf{D} = \mathbf{0}$,

(ii) $\mathbf{QKX} = \mathbf{0}$,

(iii) $\mathbf{X}^H\mathbf{KQ} = \mathbf{0}$,

(iv) $\mathbf{X}^H\mathbf{KX} = \mathbf{0}$,

(v) $\mathbf{X}^H\mathbf{MX} = \mathbf{D}$,

(vi) $\mathbf{L} \in \mathbf{X}\{1\}$, *and*

(vii) $\mathbf{XLQ} = \mathbf{D}$,

where $\mathbf{D} = \mathbf{Q} - \mathbf{QKQ}$. *Then* $\mathbf{c}^H\mathbf{Ly}$ *is the BLUE for* $\mathbf{c}^H\mathbf{h}$ *and the minimum variance attained is given by* $\mathbf{c}^H\mathbf{Mc}$.

Proof. See the proofs of Theorems 6.4.3–6.4.7 in Campbell and Meyer [31]. □

It turns out that the BLUE for $\mathbf{H}$ can be expressed in a form that is very similar to (3.7). We will use Theorem 5 to demonstrate this.

Theorem 6. *Suppose that the training symbol matrix* $\mathbf{S}$ *is of full column rank and* $\mathcal{R}(\mathbf{S}^T) \subseteq \mathcal{R}(\mathbf{Q}_N)$, *where* $\mathbf{Q}_N$ *is the time correlation matrix of the jammer. Then the BLUE for the channel matrix* $\mathbf{H}$ *is given by*

$$\hat{\mathbf{H}} = \mathbf{Y}\mathbf{Q}_N^{\dagger}\mathbf{S}^H(\mathbf{S}\mathbf{Q}_N^{\dagger}\mathbf{S}^H)^{\dagger} \tag{3.12}$$

with the minimum total variance (MSE)

$$MSE = tr(\mathbf{S}\mathbf{Q}_N^{\dagger}\mathbf{S}^H)^{\dagger} \cdot tr(\mathbf{Q}_r). \tag{3.13}$$

Proof. Let

$$\begin{aligned}\mathbf{K} &= \mathbf{Q}_N^\dagger \otimes \mathbf{Q}_r^\dagger - [\mathbf{Q}_N^\dagger \mathbf{S}^T (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger \mathbf{S}^* \mathbf{Q}_N^\dagger] \otimes \mathbf{Q}_r^\dagger \\ \mathbf{L} &= [(\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger \mathbf{S}^* \mathbf{Q}_N^\dagger] \otimes \mathbf{I}_{n_r} \\ \mathbf{M} &= (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger \otimes \mathbf{Q}_r.\end{aligned}$$

Recall that $\mathbf{Q} = \mathbf{Q}_N \otimes \mathbf{Q}_r$ and $\mathbf{X} = \mathbf{S}^T \otimes \mathbf{I}_{n_r}$. Hence, by employing the properties of the Moore-Penrose inverse defined in Definition 2, it is not hard to see that

$$\begin{aligned}\mathbf{D} &= [\mathbf{Q}_N \mathbf{Q}_N^\dagger \mathbf{S}^T (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger \mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{Q}_N] \otimes \mathbf{Q}_r^\dagger \\ &= [\mathbf{S}^T (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger \mathbf{S}^*] \otimes \mathbf{Q}_r^\dagger.\end{aligned}$$

Indeed, we note that $\mathbf{Q}_N \mathbf{Q}_N^\dagger$ is the projection operator onto $\mathcal{R}(\mathbf{Q}_N)$ [31]. Since $\mathcal{R}(\mathbf{S}^T) \subseteq \mathcal{R}(\mathbf{Q}_N)$, we have $\mathbf{Q}_N \mathbf{Q}_N^\dagger \mathbf{S}^T = \mathbf{S}^T$. Hence the second equality above results.

We are going to show that this choice of $\mathbf{K}$, $\mathbf{L}$, and $\mathbf{M}$ satisfies the seven conditions in Theorem 5.

(i) Note that $\mathcal{R}(\mathbf{S}^T(\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger) \subseteq \mathcal{R}(\mathbf{S}^T)$. Hence $\mathcal{R}(\mathbf{D}) \subseteq \mathcal{R}(\mathbf{S}^T \otimes \mathbf{I}_{n_r})$ [29]. Since $\mathbf{I}_{Nn_r} - \mathbf{X}\mathbf{X}^\dagger$ is the projection operator onto the orthogonal complement of $\mathcal{R}(\mathbf{S}^T \otimes \mathbf{I}_{n_r})$, $(\mathbf{I}_{Nn_r} - \mathbf{X}\mathbf{X}^\dagger)\mathbf{D} = \mathbf{0}$.

(ii) Like above, it is easy to work out that

$$\mathbf{Q}\mathbf{K}\mathbf{X} = [\mathbf{S}^T - \mathbf{S}^T (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)] \otimes \mathbf{Q}_r \mathbf{Q}_r^\dagger.$$

Hence, if $(\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T)^\dagger (\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T) = \mathbf{I}_{n_r}$, then $\mathbf{Q}\mathbf{K}\mathbf{X} = \mathbf{0}$. From [31], what we need is that $\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T$ is of full rank (i.e., non-singular). To see that this is indeed true, first we note that $\mathbf{Q}_N^\dagger = \tilde{\mathbf{U}}_N \tilde{\mathbf{\Lambda}}^{-1} \tilde{\mathbf{U}}_N^H$. Let $\tilde{N}$ denote the rank of $\mathbf{Q}_N$. Since $\mathcal{R}(\mathbf{S}^T) \subseteq \mathcal{R}(\mathbf{Q}_N)$ and $\mathbf{S}^T$ is of full column rank, $\tilde{N} \geq n_t$ and $\mathbf{S}^T = \tilde{\mathbf{U}}_N \tilde{\mathbf{S}}^T$, where $\tilde{\mathbf{S}}$ is a $n_t \times \tilde{N}$ matrix with full column rank. Thus, $\mathbf{S}^* \mathbf{Q}_N^\dagger \mathbf{S}^T = \tilde{\mathbf{S}}^* \tilde{\mathbf{\Lambda}}_N^{-1} \tilde{\mathbf{S}}^T$ is non-singular.

(iii) Similar to (ii).

(iv), (v) & (vii) Direct substitutions verify $\mathbf{X}^H\mathbf{K}\mathbf{X} = \mathbf{0}$, $\mathbf{X}^H\mathbf{M}\mathbf{X} = \mathbf{D}$, and $\mathbf{X}^H\mathbf{L}\mathbf{Q} = \mathbf{D}$.

(vi) Indeed, $\mathbf{X}\mathbf{L}\mathbf{X} = [\mathbf{S}^T(\mathbf{S}^*\mathbf{Q}_N^\dagger\mathbf{S}^T)^\dagger(\mathbf{S}^*\mathbf{Q}_N^\dagger\mathbf{S}^T)] \otimes \mathbf{I}_{n_r} = \mathbf{S}^T \otimes \mathbf{I}_{n_r} = \mathbf{X}$, where the second equality is due to the fact that $\mathbf{S}^*\mathbf{Q}_N^\dagger\mathbf{S}^T$ is non-singular.

By Theorem 5, for $i = 1, 2, \ldots, n_t n_r$, $\mathbf{e}_i^H\mathbf{L}\mathbf{y}$ is the BLUE for $\mathbf{e}_i^H\mathbf{h}$ with variance $\mathbf{e}_i^H\mathbf{M}\mathbf{e}_i$. Eqns. (3.12) and (3.13) are simply compact ways to express these results. To obtain (3.13), we have used the fact that $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \cdot \mathrm{tr}(\mathbf{B})$ when $\mathbf{A}$ and $\mathbf{B}$ are both square matrices [29]. □

3.2 Channel Estimation with Bayesian Estimator

3.2.1 System Model

We consider a transmitter-receiver pair with n_t transmit antennas and n_r receive antennas over a frequency flat fading channel in the presence of colored Gaussian interference. We assume that the colored interference is composed of the thermal noise and jamming signals transmitted by multiple jammers. The jth interferer has n_j transmit antennas. We assume that the transmission from the transmitter to the receiver is packetized. Each packet contains a training frame that is composed of a set of known training sequences, each of which is sent out by a transmit antenna. In matrix notation, the observed training symbols at the receiver for a packet are given by

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \underbrace{\sum_{i=1}^{M} \mathbf{H}_i\mathbf{S}_i + \mathbf{W}}_{\mathbf{E}} = \mathbf{H}\mathbf{S} + \mathbf{E}, \tag{3.14}$$

where where $\mathbf{S}$ is the $n_t \times N$ transmitted training symbol matrix that is known to the receiver, N is the number of training symbols, and $\mathbf{S}_i$ is the $n_i \times N$ jamming signal matrix from the ith jammer. We assume that symbols in $\mathbf{S}_i$ are identically distributed zero-mean, circular-symmetric complex Gaussian random variables, correlated across both space and time. We assume that the number of training symbols N is larger

than n_t. The jamming signals of the jammers are independent on each other. In addition, the jamming processes are assumed to be wide-sense stationary. We assume that the number of training symbols N is larger than n_t. The $n_r \times n_t$ matrix $\mathbf{H}$ and $n_r \times n_i$ matrix $\mathbf{H}_i$ are the channel matrices from the transmitter and the ith jammer to the receiver, respectively. We assume that the elements in $\mathbf{H}$ and $\mathbf{H}_i$ are independent, identically distributed zero mean complex Gaussian random variables with variance σ^2 and σ_i^2, respectively. In addition, $\mathbf{W}$ is an additive white Gaussian noise (AWGN) matrix and the elements in $\mathbf{W}$ are assumed to be independent, zero-mean, circular-symmetric, complex Gaussian random variables with variance σ_w^2. Finally, $\mathbf{H}, \mathbf{H}_1, \ldots, \mathbf{H}_M, \mathbf{S}_1, \ldots, \mathbf{S}_M$, and $\mathbf{W}$ are all independent of one another.

3.2.2 Bayesian Channel Estimator

Let $\mathbf{y} = vec(\mathbf{Y})$, $\mathbf{h} = vec(\mathbf{H})$, and $\mathbf{e} = vec(\mathbf{E})$, where $vec(\mathbf{X})$ is the vector obtained by stacking the columns of $\mathbf{X}$ on top of each other . Taking transpose and then vectorizing on both sides of (3.14), we have [29]

$$\mathbf{y} = (\mathbf{S}^T \otimes \mathbf{I}_{n_r})\mathbf{h} + \mathbf{e}, \tag{3.15}$$

where $\otimes$, $(\cdot)^T$, and $\mathbf{I}_{n_r}$ denote the Kronecker product, transpose of a matrix and $n_r \times n_r$ identity matrix, respectively. In (3.15), $\mathbf{h}$ is a channel vector with distribution $\mathcal{N}_c(0, \mathbf{C}_H)$ where the covariance matrix $\mathbf{C}_H = E[\mathbf{h}\mathbf{h}^H] = \sigma^2 \mathbf{I}_{n_r n_t}$ and $\mathbf{e}$ is a noise vector with distribution $\mathcal{N}_c(0, \mathbf{Q})$ where the covariance matrix $\mathbf{Q} = E[\mathbf{e}\mathbf{e}^H]$. We assume that $\mathbf{e}$ is independent of $\mathbf{h}$. During the training period, the Bayesian estimator of the channel vector $\mathbf{h}$ based on the received training block $\mathbf{Y}$, introduced in Section 2.3, can be obtained as

$$\hat{\mathbf{h}} = [(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H \mathbf{Q}^{-1} (\mathbf{S}^T \otimes \mathbf{I}_{n_r}) + \mathbf{C}_H^{-1}]^{-1} (\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H \mathbf{Q}^{-1} \mathbf{y}. \tag{3.16}$$

We note that the noise vector

$$\begin{aligned} \mathbf{e} &= \sum_{i=1}^{M} vec(\mathbf{H}_i \mathbf{S}_i) + vec(\mathbf{W}) \\ &= \sum_{i=1}^{M} (\mathbf{S}_i{}^T \otimes \mathbf{H}_i) vec(\mathbf{I}_{n_i}) + vec(\mathbf{W}). \end{aligned} \tag{3.17}$$

Let s_{kj}^i be the jamming symbol transmitted by the kth antenna of the ith jammer at at time j and

$$E[s_{kj}^i s_{lm}^{i\,*}] = \begin{cases} R_{k,k}^i(j,m) & \text{for} \quad k = l \\ R_{k,l}^i(j,m) & \text{for} \quad k \neq l, \end{cases}$$

where $R_{k,k}^i(j,m)$ is the time correlation between the jamming symbols at times j and m from the kth antenna of the ith jammer and $R_{k,l}^i(j,m)$ is the spatial correlation between the jamming symbol at time j from the kth antenna and the jamming symbol at time m from the lth antenna of the ith jammer. Then the $Nn_r \times Nn_r$ noise correlation matrix is given by

$$\begin{aligned} \mathbf{Q} &= \underbrace{\sum_{i=1}^{M} E[(\mathbf{S}_i{}^T \otimes \mathbf{H}_i) vec(\mathbf{I}_{n_i}) vec(\mathbf{I}_{n_i})^H (\mathbf{S}_i{}^T \otimes \mathbf{H}_i)^H]}_{\mathbf{J}} \\ &\quad + E[vec(\mathbf{W}) vec(\mathbf{W})^H] \\ &= \mathbf{J} + \sigma_w^2 \mathbf{I}_{Nn_r}. \end{aligned} \tag{3.18}$$

In (3.18), the $Nn_r \times Nn_r$ matrix $\mathbf{J}$ can be easily obtained as

$$\mathbf{J} = \sum_{i=1}^{M} \sigma_i^2 \begin{bmatrix} \sum_{k=1}^{n_i} R_{k,k}^i(0) \cdot \mathbf{I}_{n_r} & \cdots & \sum_{k=1}^{n_i} R_{k,k}^i(N-1) \cdot \mathbf{I}_{n_r} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R_{k,k}^i(N-1) \cdot \mathbf{I}_{n_r} & \cdots & \sum_{k=1}^{n_i} R_{k,k}^i(0) \cdot \mathbf{I}_{n_r} \end{bmatrix} \tag{3.19}$$

$$= \sum_{i=1}^{M} (\sigma_i^2 \mathbf{Q}_N^{(i)} \otimes \mathbf{I}_{n_r}), \tag{3.20}$$

where

$$\mathbf{Q}_N^{(i)} = \begin{bmatrix} \sum_{k=1}^{n_i} R_{k,k}^i(0) & \cdots & \sum_{k=1}^{n_i} R_{k,k}^i(N-1) \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R_{k,k}^i(N-1) & \cdots & \sum_{k=1}^{n_i} R_{k,k}^i(0) \end{bmatrix} \tag{3.21}$$

and $R_{k,k}^i(m-n) = R_{k,k}^i(m,n)$ because of the wide-sense stationary assumption. The derivation for $\mathbf{J}$ in (3.19) can be found in Section 3.4. Due to the i.i.d assumption we made on the elements of the channel $\mathbf{H}_i$, the space correlations between interference symbols from different antennas play no role in the correlation matrix $\mathbf{Q}$. As a result, only time correlation terms remain in (3.19).

Thus the noise correlation matrix $\mathbf{Q}$ is given by

$$\begin{aligned} \mathbf{Q} &= \sum_{i=1}^{M} (\sigma_i^2 \mathbf{Q}_N^{(i)} \otimes \mathbf{I}_{n_r}) + \sigma_w^2 \mathbf{I}_{Nn_r} \\ &= (\sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N) \otimes \mathbf{I}_{n_r}. \end{aligned} \tag{3.22}$$

The second equality is obtained from $(\mathbf{A} \otimes \mathbf{B}) + (\mathbf{C} \otimes \mathbf{B}) = (\mathbf{A} + \mathbf{C}) \otimes \mathbf{B}$ [29]. By using the Kronecker product form of $\mathbf{Q}$ obtained in (3.22), the Bayesian channel estimator in (3.16) can be reduced to

$$\hat{\mathbf{h}} = \left\{ \left[\left(\mathbf{S}^* \mathbf{A}_N^{-1} \mathbf{S}^T + \frac{1}{\sigma_H^2} \mathbf{I}_{n_t} \right)^{-1} \mathbf{S}^* \mathbf{A}_N^{-1} \right] \otimes \mathbf{I}_{n_r} \right\} \mathbf{y}, \tag{3.23}$$

where

$$\mathbf{A}_N = \sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N. \tag{3.24}$$

In addition, we can write (3.23) back into matrix form as

$$\hat{\mathbf{H}} = \mathbf{Y} \mathbf{A}_N^{-1} \mathbf{S}^H \left(\mathbf{S} \mathbf{A}_N^{-1} \mathbf{S}^H + \frac{1}{\sigma^2} \mathbf{I}_{n_t} \right)^{-1}. \tag{3.25}$$

Moreover the MSE of the Bayesian estimator for $\mathbf{H}$, discussed in Section 2.3, is given by

$$\begin{aligned}
\mathrm{MSE}^{(N)} &= \mathrm{tr}[(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H \mathbf{Q}^{-1} (\mathbf{S}^T \otimes \mathbf{I}_{n_r}) + \mathbf{C}_H^{-1}]^{-1} \\
&= \mathrm{tr}[(\mathbf{S}^T \otimes \mathbf{I}_{n_r})^H \{\sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N) \otimes \mathbf{I}_{n_r}\}^{-1} (\mathbf{S}^T \otimes \mathbf{I}_{n_r}) \\
&\quad + (\sigma^2 \mathbf{I}_{n_r n_t})^{-1}]^{-1} \\
&= \mathrm{tr}[\{\mathbf{S}^* (\sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N)^{-1} \mathbf{S}^T \otimes \mathbf{I}_{n_r}\} + \{\frac{1}{\sigma^2} \mathbf{I}_{n_t} \otimes \mathbf{I}_{n_r}\}]^{-1} \\
&= \mathrm{tr} \left[\left\{ \mathbf{S}^* (\sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N)^{-1} \mathbf{S}^T + \frac{1}{\sigma^2} \mathbf{I}_{n_t} \right\} \otimes \mathbf{I}_{n_r} \right]^{-1} \\
&= \mathrm{tr}[\{\underbrace{\mathbf{S}^* (\sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^{(i)} + \sigma_w^2 \mathbf{I}_N)^{-1} \mathbf{S}^T}_{\mathbf{A}_N} + \frac{1}{\sigma^2} \mathbf{I}_{n_t}\}^{-1}] \cdot n_r \\
&= n_r \cdot \mathrm{tr} \left[\left\{ \mathbf{S}^* \mathbf{A}_N^{-1} \mathbf{S}^T + \frac{1}{\sigma^2} \mathbf{I}_{n_t} \right\}^{-1} \right], \qquad (3.26)
\end{aligned}$$

where $(\cdot)^*$, $(\cdot)^H$, and $\mathrm{tr}(\cdot)$ denote the complex-conjugate, complex-conjugate (Hermitian) transpose and trace of a matrix, respectively. The third and fourth equalities are obtained by using $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, $(\mathbf{A} \otimes \mathbf{B}) + (\mathbf{C} \otimes \mathbf{B}) = (\mathbf{A} + \mathbf{C}) \otimes \mathbf{B}$, and $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$. The fifth equality is due to $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$ [29].

We assume that the channel matrices $\mathbf{H}$ and $\mathbf{H}_i$ for $i = 1, \ldots, M$ are short-term statistics that may change from packet to packet. On the other hand, the interference correlation matrix $\mathbf{Q}$ varies at a rate that is much slower than that of the channel matrices. As a result, it is possible for the receiver to estimate $\mathbf{Q}$ using a number of previous packets and feed back relevant information to the transmitter so that it can make use of this information to select the optimal training sequence set for the estimation of $\mathbf{H}$ during the current packet.

3.3 Summary

We described a transmitter-receiver pair in MIMO system model over frequency flat fading channel in the presence of colored interference. We developed the BLUE for the MIMO channel and derived the MSE expressions of the BLUE when jammer covariance matrices are non-singular and singular cases. In the Bayesian approach, we considered the case where there are multiple interferers. We obtained the Bayesian estimator and derived the MSE expression of the estimator. From the expression of the MSE for both estimators, we note that the MSE of the channel estimator depends on the choice of the training sequence set $\mathbf{S}$. We will discuss how to optimize the training sequences to minimize MSE of the channel estimator in the following chapter.

3.4 Derivation of the J Matrix

Let the $n_i \times N$ jamming signal matrix $\mathbf{S}_i$ and $n_r \times n_i$ channel matrix from the jammer $\mathbf{H}_i$ be

$$\mathbf{S}_i = \begin{bmatrix} s^i_{1,1} & \cdots & s^i_{1,N} \\ \vdots & & \vdots \\ s^i_{n_i,1} & \cdots & s^i_{n_i,N} \end{bmatrix}$$

and

$$\mathbf{H}_i = \begin{bmatrix} h^i_{1,1} & \cdots & h^i_{1,n_i} \\ \vdots & & \vdots \\ h^i_{n_r,1} & \cdots & h^i_{n_r,n_i} \end{bmatrix}.$$

We note from (3.17) that noise vector $\mathbf{e}$ is expressed by

$$\mathbf{e} = \sum_{i=1}^{M} (\mathbf{S}_i{}^T \otimes \mathbf{H}_i) vec(\mathbf{I}_{n_i}) + vec(\mathbf{N}). \tag{3.27}$$

Let us just consider the first term in (3.27)and let $\hat{\mathbf{e}}$ be

$$\begin{aligned}
\hat{\mathbf{e}} &= \sum_{i=1}^{M} ({\mathbf{S}_i}^T \otimes \mathbf{H}_i) vec(\mathbf{I}_{n_i}) \\
&= \sum_{i=1}^{m} \left(\begin{bmatrix} s^i_{1,1} & s^i_{2,1} & \cdots & s^i_{n_i,1} \\ \vdots & \vdots & & \vdots \\ s^i_{1,N} & s^i_{2,N} & \cdots & s^i_{n_i,N} \end{bmatrix} \otimes \mathbf{H}_i \right) vec(\mathbf{I}_{n_i}) \\
&= \sum_{i=1}^{m} \underbrace{\begin{bmatrix} s^i_{1,1}\mathbf{H}_i & s^i_{2,1}\mathbf{H}_i & \cdots & s^i_{n_i,1}\mathbf{H}_i \\ \vdots & \vdots & & \vdots \\ s^i_{1,N}\mathbf{H}_i & s^i_{2,N}\mathbf{H}_i & \cdots & s^i_{n_i,N}\mathbf{H}_i \end{bmatrix}}_{N \cdot n_r \times n_i \cdot n_i} \underbrace{vec(\mathbf{I}_{n_i})}_{n_i \cdot n_i \times 1}
\end{aligned} \tag{3.28}$$

We can easily see that $[vec(\mathbf{I}_{n_i})]^T$ is

$$[vec(\mathbf{I}_{n_i})]^T = \underbrace{[\underbrace{1\ 0\ \cdots\ 0}_{n_i} \underbrace{0\ 1\ 0\ \cdots\ 0}_{n_i} \underbrace{0\ 0\ 1\ 0\ \cdots\ 0}_{n_i} \cdots\cdots\cdots \underbrace{0\ 0\ \cdots\ 0\ 1}_{n_i}]}_{1 \times n_i \cdot n_i}. \tag{3.29}$$

Therefore we can obtain (3.28) by simple matrix multiplication and (3.28) can be rewritten by the following $N \cdot n_r \times 1$ vector $\hat{\mathbf{e}}$

$$\hat{\mathbf{e}} = \sum_{i=1}^{M} \left[\begin{array}{l} \left. \begin{array}{l} s^i_{1,1}h^i_{1,1} + s^i_{2,1}h^i_{1,2} + s^i_{3,1}h^i_{1,3} + \cdots + s^i_{n_i,1}h^i_{1,n_i} \\ s^i_{1,1}h^i_{2,1} + s^i_{2,1}h^i_{2,2} + s^i_{3,1}h^i_{2,3} + \cdots + s^i_{n_i,1}h^i_{2,n_i} \\ \vdots \\ s^i_{1,1}h^i_{n_r,1} + s^i_{2,1}h^i_{n_r,2} + s^i_{3,1}h^i_{n_r,3} + \cdots + s^i_{n_i,1}h^i_{n_r,n_i} \end{array} \right\} n_r \times 1 \\ \left. \begin{array}{l} s^i_{1,2}h^i_{1,1} + s^i_{2,2}h^i_{1,2} + s^i_{3,2}h^i_{1,3} + \cdots + s^i_{n_i,2}h^i_{1,n_i} \\ s^i_{1,2}h^i_{2,1} + s^i_{2,2}h^i_{2,2} + s^i_{3,2}h^i_{2,3} + \cdots + s^i_{n_i,2}h^i_{2,n_i} \\ \vdots \\ s^i_{1,2}h^i_{n_r,1} + s^i_{2,2}h^i_{n_r,2} + s^i_{3,2}h^i_{n_r,3} + \cdots + s^i_{n_i,2}h^i_{n_r,n_i} \end{array} \right\} n_r \times 1 \\ \vdots \\ \vdots \\ \vdots \\ \left. \begin{array}{l} s^i_{1,N}h^i_{1,1} + s^i_{2,N}h^i_{1,2} + s^i_{3,N}h^i_{1,3} + \cdots + s^i_{n_i,N}h^i_{1,n_i} \\ s^i_{1,N}h^i_{2,1} + s^i_{2,N}h^i_{2,2} + s^i_{3,N}h^i_{2,3} + \cdots + s^i_{n_i,N}h^i_{2,n_i} \\ \vdots \\ s^i_{1,N}h^i_{n_r,1} + s^i_{2,N}h^i_{n_r,2} + s^i_{3,N}h^i_{n_r,3} + \cdots + s^i_{n_i,N}h^i_{n_r,n_i} \end{array} \right\} n_r \times 1 \end{array} \right\} N \cdot n_r \times 1$$

Let the $N \cdot n_r \times N \cdot n_r$ covariance matrix of $\hat{\mathbf{e}}$ be $\mathbf{K} = E[\hat{\mathbf{e}}\hat{\mathbf{e}}^H]$. By using the assumption on the jamming signal matrix and the i.i.d assumption on the elements of the $\mathbf{H}_i$, the covariance matrix $\mathbf{K}$ can be reduced to as

$$\mathbf{K} = \sum_{i=1}^{M} \begin{bmatrix} \hat{\mathbf{K}}^i(0) & \cdots & \hat{\mathbf{K}}^i(N-1) \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{K}}^i(N-1) & \cdots & \hat{\mathbf{K}}^i(0) \end{bmatrix}, \tag{3.30}$$

where τ in $\hat{\mathbf{K}}^i(\tau)$ is the time difference between jamming symbols for $\tau = 0, 1, 2, \cdots, N-1$, $\hat{\mathbf{K}}^i(\tau)$ is an $n_r \times n_r$ diagonal matrix and is given by

$$\hat{\mathbf{K}}^i(\tau) = \begin{bmatrix} \sum_{k=1}^{n_i} R^i_{k,k}(\tau) \cdot \sigma_i^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{k=1}^{n_i} R^i_{k,k}(\tau) \cdot \sigma_i^2 \end{bmatrix}, \tag{3.31}$$

where n_i is the number of antennas at the ith interferer, $R^i_{k,k}(\tau)$ is the time correlation between the jamming symbols from the kth antenna of the ith jammer and σ_i^2 is the variance of the channel from the jammer. All the space correlation terms are removed due to the i.i.d assumption on elements of the channel from the jammer. Moreover, (3.31) can be further reduced to

$$\hat{\mathbf{K}}^i(\tau) = \sum_{k=1}^{n_i} R^i_{k,k}(\tau) \cdot \sigma_i^2 \cdot \mathbf{I}_{n_r}. \tag{3.32}$$

Thus, the covariance matrix $\mathbf{K}$ is expressed by

$$\begin{aligned} \mathbf{K} &= \sum_{i=1}^{M} \sigma_i^2 \begin{bmatrix} \sum_{k=1}^{n_i} R^i_{k,k}(0) \cdot \mathbf{I}_{n_r} & \cdots & \sum_{k=1}^{n_i} R^i_{k,k}(N-1) \cdot \mathbf{I}_{n_r} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R^i_{k,k}(N-1) \cdot \mathbf{I}_{n_r} & \cdots & \sum_{k=1}^{n_i} R^i_{k,k}(0) \cdot \mathbf{I}_{n_r} \end{bmatrix} \\ &= \sum_{i=1}^{M} \sigma_i^2 \begin{bmatrix} \sum_{k=1}^{n_i} R^i_{k,k}(0) & \cdots & \sum_{k=1}^{n_i} R^i_{k,k}(N-1) \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n_i} R^i_{k,k}(N-1) & \cdots & \sum_{k=1}^{n_i} R^i_{k,k}(0) \end{bmatrix} \otimes \mathbf{I}_{n_r}. \end{aligned} \tag{3.33}$$

Therefore we finally obtain $\mathbf{J}$ in (3.19).

Table 3–1: Matrix Notation

$\mathbf{A}$	capital letters in boldface denote matrices
$\mathbf{a}$	lowercase letters in boldface denote column vectors
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{0}$	zero matrix
$\text{diag}(x_1, \ldots, x_n)$	diagonal matrix with $x_1, \ldots, x_n$ as the diagonal elements
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of $\mathbf{A}$
$\mathbf{A}^H$	complex conjugate transpose (Hermitian) of $\mathbf{A}$
$\text{tr}(\mathbf{A})$	trace of $\mathbf{A}$
$\text{rank}(\mathbf{A})$	rank of $\mathbf{A}$
$\mathbf{A}^-$	(1)-inverse of $\mathbf{A}$
$\mathbf{A}^\dagger$	Moore-Penrose inverse of $\mathbf{A}$
$\text{vec}(\mathbf{A})$	vector obtained by stacking columns of $\mathbf{A}$ on top of each other
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\mathcal{R}(\mathbf{A})$	range of $\mathbf{A}$
$\mathcal{N}(\mathbf{A})$	null space of $\mathbf{A}$

CHAPTER 4
TRAINING SEQUENCE OPTIMIZATION

We note from Chapter 3 that the MSE of the channel estimator depends on the choice of the training sequence set $\mathbf{S}$. Hence, it is natural to ask whether there is an optimal set of training sequences that gives the best estimation performance. Moreover, it is conceivable that the optimal training sequence set will depend on the characteristic of the interference at the receiver. In this chapter, we determine the optimal training sequence set that can minimize the MSE of the channel estimator under a total transmit power constraint.

4.1 Optimal Training Sequence Set with BLUE

We note from (3.8) and (3.13) in Chapter 3 that MSE of the BLUE channel estimator depends on the choice of the training sequence set. In this section, we determine optimal training sequence set when we employ BLUE channel estimator. To have a meaningful formulation of the sequence optimization problem, we need to consider the following two restrictions. First, we limit the maximum total transmit power of transmit antenna array to P. Second, we assume that the time correlation matrix, $\mathbf{Q}_N$, of the jammer is non-singular. In addition, we recall that $\mathbf{S}^T$ has to be of full column rank for $\mathbf{H}$ to be linearly unbiasedly estimable. The following theorem provides a constructive method to obtain the optimal training sequence set under these restrictions.

Theorem 7. *Suppose that the time correlation matrix, $\mathbf{Q}_N$, of the jammer is non-singular. Let $\mathbf{V}$ be an arbitrary $n_t \times n_t$ unitary matrix. Also let $\lambda_1^{(N)}, \ldots, \lambda_{n_t}^{(N)}$ be the n_t smallest eigenvalues of $\mathbf{Q}_N$, and $\tilde{\mathbf{U}}_N$ is the $N \times n_t$ matrix whose columns are the eigenvectors of $\mathbf{Q}_N$ corresponding to $\lambda_1^{(N)}, \ldots, \lambda_{n_t}^{(N)}$, respectively. Then the training*

sequence set

$$\mathbf{S}_* = \mathbf{V}\ diag\left(\sqrt{\frac{NP\sqrt{\lambda_1^{(N)}}}{\sum_{j=1}^{n_t}\sqrt{\lambda_j^{(N)}}}},\ldots,\sqrt{\frac{NP\sqrt{\lambda_{n_t}^{(N)}}}{\sum_{j=1}^{n_t}\sqrt{\lambda_j^{(N)}}}}\right)\bar{\mathbf{U}}_N^H \tag{4.1}$$

gives the smallest mean square estimation error of

$$MSE_*^{(N)} = \frac{tr(\mathbf{Q}_r)}{NP}\left[\sum_{j=1}^{n_t}\sqrt{\lambda_j^{(N)}}\right]^2 \tag{4.2}$$

among all sequence sets that satisfy $tr(\mathbf{S}\mathbf{S}^H) \leq NP$ *and* $\mathbf{S}^T$ *is of full column rank, when the corresponding BLUE is employed to estimate* $\mathbf{H}$*. Moreover, the BLUE for* $\mathbf{H}$ *becomes*

$$\hat{\mathbf{H}}_* = \mathbf{Y}\bar{\mathbf{U}}_N\ diag\left(\sqrt{\frac{\sum_{j=1}^{n_t}\sqrt{\lambda_j^{(N)}}}{NP\sqrt{\lambda_1^{(N)}}}},\ldots,\sqrt{\frac{\sum_{j=1}^{n_t}\sqrt{\lambda_j^{(N)}}}{NP\sqrt{\lambda_{n_t}^{(N)}}}}\right)\mathbf{V}^H = \mathbf{Y}\mathbf{S}_*^{\dagger} \tag{4.3}$$

when the optimal sequence set $\mathbf{S}_*$ *is employed.*

Proof. We note that the MSE of the BLUE for $\mathbf{H}$ depends on the training sequence set $\mathbf{S}$ only through the first term on the right hand side of (3.13). Moreover, the Moore-Penrose inverse reduces to the usual matrix inverse for a non-singular square matrix. Thus, it suffices to consider the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{S}}\ tr(\mathbf{S}\mathbf{Q}_N^{-1}\mathbf{S}^H)^{-1} \\ &\text{subject to } \operatorname{rank}(\mathbf{S}) = n_t \text{ and } tr(\mathbf{S}\mathbf{S}^H) \leq NP. \end{aligned} \tag{4.4}$$

Let $\bar{\mathbf{S}} = \mathbf{S}\mathbf{Q}_N^{-1/2}$. We can rewrite the optimization problem in (4.4) into the following form:

$$\begin{aligned} &\min_{\bar{\mathbf{S}}}\ tr(\bar{\mathbf{S}}\bar{\mathbf{S}}^H)^{-1} \\ &\text{subject to } \operatorname{rank}(\bar{\mathbf{S}}) = n_t \text{ and } tr(\bar{\mathbf{S}}\mathbf{Q}_N\bar{\mathbf{S}}^H) \leq NP. \end{aligned} \tag{4.5}$$

Further, let $\mu_1, \ldots, \mu_{n_t}$ be the (positive) eigenvalues of $\bar{\mathbf{S}}\bar{\mathbf{S}}^H$ arranged in a descending order and $\lambda_1^{(N)}, \ldots, \lambda_N^{(N)}$ be the eigenvalues of $\mathbf{Q}_N$ arranged in an ascending order. To proceed, we need to make use of the following result, whose proof can be found, for example, in [32, pp. 249].

Lemma 1. *Suppose that* $\mathbf{X}$ *and* $\mathbf{Y}$ *are two Hermitian* $N \times N$ *matrices. Arrange the eigenvalues* $x_i, \ldots, x_N$ *of* $\mathbf{X}$ *in a descending order and the eigenvalues* $y_i, \ldots, y_N$ *of* $\mathbf{Y}$ *in an ascending order. Then* $tr(\mathbf{XY}) \geq \sum_{i=1}^{N} x_i y_i$.

Applying this lemma to the constraint in (4.5), we can bound

$$\mathrm{tr}(\bar{\mathbf{S}}\mathbf{Q}_N\bar{\mathbf{S}}^H) = \mathrm{tr}(\mathbf{Q}_N\bar{\mathbf{S}}^H\bar{\mathbf{S}}) \geq \sum_{i=1}^{n_t} \mu_i \lambda_i^{(N)}. \tag{4.6}$$

Now, consider the following relaxed optimization problem:

$$\min_{\mu_1,\ldots,\mu_{n_t}} \sum_{i=1}^{n_t} \frac{1}{\mu_i}$$
$$\text{subject to } \sum_{i=1}^{n_t} \mu_i \lambda_i^{(N)} \leq NP \text{ and } \mu_1 \geq \mu_2 \geq \cdots \geq \mu_{n_t} > 0. \tag{4.7}$$

If we can construct a matrix $\bar{\mathbf{S}}$ such that the eigenvalues of $\bar{\mathbf{S}}\bar{\mathbf{S}}^H$ are exactly the solution of the above relaxed optimization problem and that $\mathrm{tr}(\bar{\mathbf{S}}\mathbf{Q}_N\bar{\mathbf{S}}^H\} = \sum_{i=1}^{n_t} \mu_i \lambda_i^{(N)}$, then this choice of $\bar{\mathbf{S}}$ will be a solution of the original sequence optimization problem in (4.5).

The relaxed optimization problem (4.7) can be solved by the standard Karush-Kuhn-Tucker condition technique [33] since the cost function and the constraint are both convex. The optimal solution is given by

$$\mu_i^* = \frac{NP}{\sum_{j=1}^{n_t} \sqrt{\lambda_j^{(N)}}} \cdot \frac{1}{\sqrt{\lambda_i^{(N)}}}, \tag{4.8}$$

for $i = 1, 2, \ldots, n_t$. The derivation of this solution is given in Section 4.4. In addition, the lower bound in (4.11) can be achieved by choosing $\bar{\mathbf{S}} = \mathbf{V}\mathrm{diag}(\sqrt{\mu_1^*}, \ldots, \sqrt{\mu_{n_t}^*})\bar{\mathbf{U}}_N^H$, where $\mathbf{V}$ and $\bar{\mathbf{U}}_N$ are as specified in the statement of the theorem, and $\mu_1, \ldots, \mu_{n_t}$

are given in (4.8). Hence, we get (4.1) by transforming back from the optimal choice of $\bar{\mathbf{S}}$. In addition, the minimum value of MSE in (4.2) can be obtained by plugging the optimal choice of $\mu_1^*, \ldots, \mu_{n_t}^*$ in (4.8) back into the cost function in (4.7). □

We note that not only does the choice of optimal sequence set minimize the estimation error, but this choice also simplifies the implementation complexity of the BLUE for $\mathbf{H}$. It is easy to see from (4.3) that the complexity of the BLUE with the optimal sequence set is $\mathcal{O}(n_t n_r N)$. Finally, we point out that the training sequence optimization problem is not very meaningful when $\mathbf{Q}_N$ is singular. In particular, when $\text{rank}(\mathbf{Q}_N) \leq N - n_t$, then according to the discussion in Section 3.1.2, one can obtain zero MSE, for any arbitrarily small transmit power, by simply selecting any full-rank $\mathbf{S}$ such that $\mathcal{R}(\mathbf{S}^T) \subseteq \mathcal{N}(\mathbf{Q}_N)$. When $N - n_t < \text{rank}(\mathbf{Q}_N) < N$, one can select $N - \text{rank}(\mathbf{Q}_N)$ linearly independent sequences with arbitrarily small, but non-zero, power from $\mathcal{N}(\mathbf{Q}_N)$. Then the remaining power is allocated to sequences that are found as in Theorem 7 restricting to $\mathcal{R}(\mathbf{Q}_N)$.

4.2 Optimal Training Sequence Set with Bayesian Estimator

Our goal is to minimize the MSE of the Bayesian estimator by selecting the optimal training sequence set $\mathbf{S}$ with the total energy constraint $\text{tr}(\mathbf{S}\mathbf{S}^H) \leq NP$. Therefore we can express the training optimization problem as follows:

$$\begin{aligned} \min_{\mathbf{S}} \quad & \text{tr}[\mathbf{S}^* \mathbf{A}_N^{-1} \mathbf{S}^T + \tfrac{1}{\sigma^2} \mathbf{I}_{n_t}]^{-1} \\ \text{subject to} \quad & \text{tr}(\mathbf{S}\mathbf{S}^H) \leq NP, \end{aligned} \tag{4.9}$$

where $\mathbf{A}_N = \sum_{i=1}^{M} \sigma_i^2 \mathbf{Q}_N^i + \sigma_w^2 \mathbf{I}_N$ given in (3.24). Note that with the Bayesian approach, contrary to the classical general linear model, the training symbol matrix $\mathbf{S}$ need not be of full rank to ensure the invertibility of $[\mathbf{S}^* \mathbf{A}_N^{-1} \mathbf{S}^T + \frac{1}{\sigma^2} \mathbf{I}_{n_t}]$ [25].

Let $\bar{\mathbf{S}} = \mathbf{S}\mathbf{A}_N^{-1/2}$. We can rewrite the optimization problem in (4.9) into the following

form:

$$\min_{\bar{\mathbf{S}}} \ \mathrm{tr}[\bar{\mathbf{S}}\bar{\mathbf{S}}^H + \tfrac{1}{\sigma^2}\mathbf{I}_{n_t}]^{-1} \tag{4.10}$$
$$\text{subject to} \quad \mathrm{tr}(\bar{\mathbf{S}}\mathbf{A}_N\bar{\mathbf{S}}^H) \leq NP.$$

Further, let $\mu_1, \ldots, \mu_{n_t}$ be the (positive) eigenvalues of $\bar{\mathbf{S}}\bar{\mathbf{S}}^H$ arranged in a descending order and $\lambda_1^{(N)}, \ldots, \lambda_N^{(N)}$ be the eigenvalues of $\mathbf{A}_N$ arranged in an ascending order. To proceed, we need to use *Lemma 1* in Section 4.1. Applying this lemma to the constraint in (4.10), we can bound

$$\mathrm{tr}(\bar{\mathbf{S}}\mathbf{A}_N\bar{\mathbf{S}}^H) = \mathrm{tr}(\mathbf{A}_N\bar{\mathbf{S}}^H\bar{\mathbf{S}}) \geq \sum_{i=1}^{n_t} \mu_i\lambda_i^{(N)}. \tag{4.11}$$

Now, consider the following relaxed optimization problem:

$$\min_{\mu_1,\ldots,\mu_{n_t}} \ \sum_{i=1}^{n_t} \left(\mu_i + \tfrac{1}{\sigma^2}\right)^{-1} \tag{4.12}$$
$$\text{subject to } \sum_{i=1}^{n_t} \mu_i\lambda_i^{(N)} \leq NP \text{ and } \mu_1 \geq \cdots \geq \mu_{n_t} \geq 0.$$

This relaxed optimization problem can be solved by the standard Karush-Kuhn-Tucker condition technique [33] since the cost function and the constraint are both convex.

Let

$$n_* = max\left\{ k \in \{1, 2, \ldots, n_t\} : \sqrt{\lambda_k^{(N)}} < \left(\frac{\sigma^2 NP + \sum_{i=1}^{k} \lambda_i^{(N)}}{\sum_{i=1}^{k} \sqrt{\lambda_i^{(N)}}} \right) \right\}. \tag{4.13}$$

We note that n_* above is well defined and that the inequality that defines n_* in (4.13) holds for $k = 1, \ldots, n_*$, while it does not hold for $k = n_* + 1, \ldots, n_t$. With this

definition, the optimal solution is given by

$$\mu_k^* = \begin{cases} \left(\frac{NP + \frac{1}{\sigma^2} \sum_{i=1}^{n_*} \lambda_i^{(N)}}{\sum_{i=1}^{n_*} \sqrt{\lambda_i^{(N)}}} \right) \cdot \left(\frac{1}{\sqrt{\lambda_k^{(N)}}} \right) - \frac{1}{\sigma^2} & \text{for} \quad k = 1, \ldots, n_* \\ 0 & \text{for} \quad k = n_* + 1, \ldots, n_t. \end{cases}$$

We note that this solution has the standard water-filling [34] interpretation.

If we can construct a matrix $\bar{\mathbf{S}}$ such that the eigenvalues of $\bar{\mathbf{S}}\bar{\mathbf{S}}^H$ are exactly the solution of the above relaxed optimization problem and that $\mathrm{tr}(\bar{\mathbf{S}}\mathbf{A}_N\bar{\mathbf{S}}^H) = \sum_{i=1}^{n_t} \mu_i^* \lambda_i^{(N)}$, then this choice of $\bar{\mathbf{S}}$ will be a solution of the original sequence optimization problem in (4.10). Following the same procedure described in previous section, it is easy to see that this can be done and the optimal training sequence set can then be obtained as

$$\mathbf{S}_* = \mathbf{V}\ \mathrm{diag}\left[\sqrt{\mu_1^* \lambda_1^{(N)}}, \ldots, \sqrt{\mu_{n_t}^* \lambda_{n_t}^{(N)}} \right] \bar{\mathbf{U}}_N^H, \tag{4.14}$$

where $\mathbf{V}$ is an arbitrary $n_t \times n_t$ unitary matrix and $\bar{\mathbf{U}}_N$ is the $N \times n_t$ matrix whose columns are the eigenvectors of $\mathbf{A}_N$ corresponding to the n_t smallest eigenvalues of $\mathbf{A}_N$. This optimal training sequence gives the smallest mean square estimation error of

$$\mathrm{MSE}_*^{(N)} = \frac{n_r \left(\sum_{j=1}^{n_*} \sqrt{\lambda_j^{(N)}} \right)^2}{NP + \frac{1}{\sigma^2} \sum_{j=1}^{n_*} \lambda_j^{(N)}} + n_r(n_t - n_*)\sigma^2 \tag{4.15}$$

by plugging the optimal choice of $\mu_1^*, \ldots, \mu_{n_t}^*$ back into the cost function in (4.12).

We note that not only does the choice of optimal sequence set minimize the estimation error, but this choice also simplifies the implementation complexity of the channel estimator for $\mathbf{H}$. It is easy to see that the Bayesian channel estimator in (3.25) reduces to

$$\hat{\mathbf{H}} = \mathbf{Y}\bar{\mathbf{U}}_N \mathrm{diag}[\omega_1, \ldots, \omega_{n_t}]\mathbf{V}^H, \tag{4.16}$$

where

$$\omega_k = \begin{cases} \frac{1+\sigma^2\mu_k^*}{\sqrt{\mu_k^*\lambda_1^{(N)}}} & \text{for } k = 1, \ldots, n_* \\ 0 & \text{for } k = n_* + 1, \ldots, n_t. \end{cases} \tag{4.17}$$

Thus the complexity of the Bayesian channel estimator with the optimal sequence set reduces to $\mathcal{O}(n_t n_r N)$.

4.3 Conclusion

We showed that the optimal training sequence sets can be obtained by solving two optimization problems under the total transmit power constraint for both the BLUE and Bayesian estimator cases. A simple physical interpretation of this solution is that the optimal training sequence set put its power to where the effect of the jammer is smallest, and hence the estimation error can be minimized. We note that the optimal training sequence set is an orthogonal set if $\mathbf{V}$ is chosen to be an identity matrix. However, the optimal training sequence set, in general, is not necessarily orthogonal. For instance, it is possible to obtain a choice of $\mathbf{V}$ which spreads power evenly across the transmit antennas with the use of non-orthogonal sequences. To do so, we need to construct an unitary $\mathbf{V}$ to make the diagonal elements of

$$\mathbf{S}_*\mathbf{S}_*^H = \mathbf{V}\ \mathrm{diag}\left[\mu_1^*\lambda_1^{(N)}, \ldots, \mu_{n_t}^*\lambda_{n_t}^{(N)}\right]\ \mathbf{V}^H \tag{4.18}$$

the same in both approaches. This is shown to be possible in Wong and Lok [35], and such a $\mathbf{V}$ can be constructed using a simple iterative procedure.

4.4 Solution of the Optimization Problem

In this section, the optimal solution for the training sequence optimization problem in the BLUE case in Section 4.1 is derived. To solve this optimization problem, we use the standard Karush-Kuhn-Tucker condition technique [33]. We can easily obtain the optimal solution for the Bayesian estimator case in the same way as the following derivation.

In this optimization problem, we only have inequality constraints. Thus optimization problem can be expressed as

$$\min_{\mu_1,\cdots,\mu_{N_t}} \sum_{i=1}^{N_t} \frac{1}{\mu_i} \tag{4.19}$$

$$\begin{aligned}
\text{subject to} \quad g_1(\mu) &= \sum_{i=1}^{N_t} \mu_i \lambda_i^{(N)} - NP \leq 0 \\
g_2(\mu) &= -\mu_1 \leq 0 \\
g_3(\mu) &= \mu_1 - \mu_2 \leq 0 \\
&\vdots \\
g_{N_t+1}(\mu) &= \mu_{N_t-1} - \mu_{N_t} \leq 0.
\end{aligned}$$

Thus, we need to consider three conditions given by

$$\delta \geq 0 \tag{4.20}$$

$$Df(\lambda) + \gamma^T Dh(\lambda) + \delta^T Dg(\lambda) = 0^T \tag{4.21}$$

$$\delta^T g(\lambda) = 0, \tag{4.22}$$

where δ is the KKT multiplier vector and γ is the Lagrange multiplier vector. Because we do not have any equality constraints, the $\gamma^T Dh(x)$ term in (4.21) is zero. From (4.20) and (4.21), we have

$$\delta \geq 0, \quad \delta^T = [\delta_1, \delta_2, \cdots, \delta_{N_t+1}] \tag{4.23}$$

$$f(\mu) = \sum_{i=1}^{N_t} \frac{1}{\mu_i} \tag{4.24}$$

$$Df(\mu) = [-\frac{1}{\mu_1^2}, -\frac{1}{\mu_2^2}, \cdots, -\frac{1}{\mu_{Nt}^2}] \tag{4.25}$$

$$g(\mu) = \begin{bmatrix} \sum_{i=1}^{N_t} \mu_i \lambda_i^{(N)} - NP \\ -\mu_1 \\ \mu_1 - \mu_2 \\ \vdots \\ \mu_{N_t-1} - \mu_{N_t} \end{bmatrix} \tag{4.26}$$

$$\begin{aligned} Dg(\mu) &= \begin{bmatrix} \frac{\partial g_1(\mu)}{\partial \mu_1} & \frac{\partial g_1(\mu)}{\partial \mu_2} & \cdots & \frac{\partial g_1(\mu)}{\partial \mu_{N_t}} \\ \frac{\partial g_2(\mu)}{\partial \mu_1} & \frac{\partial g_2(\mu)}{\partial \mu_2} & \cdots & \frac{\partial g_2(\mu)}{\partial \mu_{N_t}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_{N_t+1}(\mu)}{\partial \mu_1} & \frac{\partial g_{N_t+1}(\mu)}{\partial \mu_2} & \cdots & \frac{\partial g_{N_t+1}(\mu)}{\partial \mu_{N_t}} \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1^{(N)} & \lambda_2^{(N)} & \cdots & \lambda_N^{(N)} \\ -1 & 0 & \cdots & 0 \\ 1 & -1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix} \end{aligned} \tag{4.27}$$

Let $Df(\mu) + \delta^T Dg(\mu) = 0$, we have

$$\sum_{i=1}^{N_t} \frac{1}{\mu_i^2} = \delta_1 \sum_{i=1}^{Nt} \lambda_i^{(N)} - \delta_2 \tag{4.28}$$

and by (4.22), we have

$$\begin{aligned} \delta^T g(\mu) &= \delta_1 (\sum_{i=1}^{N_t} \mu_i \lambda_i^{(N)} - NP) - \delta_2 \mu_1 + \delta_3 (\mu_1 - \mu_2) \\ &+ \cdots + \delta_{N_t+1} (\mu_{N_t-1} - \mu_{N_t}) = 0. \end{aligned} \tag{4.29}$$

First, we know that all the constraints $g_i(\mu) \leq 0$. Thus in (4.29), if $g_i(\mu) < 0$, all δ_i must be zero to satisfy the equality. If $g_i(\mu) = 0$, δ_i can be either zero or positive. As a result, we note that δ_1 is the only value that we can choose as positive. Thus we have δ_i as

$$\begin{cases} \delta_i = 0, & i = 2, 3, \ldots, N_t + 1 \\ \delta_i > 0, & i = 1 \end{cases}$$

We also have $\sum_{i=1}^{N_t} \lambda_i^{(N)} \mu_i = NP$. From (4.28), we have

$$\sum_{i=1}^{N_t} (-\frac{1}{\mu_i^2} + \delta_1 \lambda_i^{(N)}) = \delta_2. \tag{4.30}$$

Because $\delta_2 = 0$, we can have expression for μ_i as

$$\mu_i = \frac{1}{\sqrt{\mu_1 C_i^{(N)}}}, \tag{4.31}$$

where $\mu_i \geq 0$. Plug (4.31) into $\sum_{i=1}^{N_t} \lambda_i^{(N)} \mu_i = NP$, we have

$$\sum_{i=1}^{N_t} \lambda_i^{(N)} \frac{1}{\sqrt{\delta_1 C_i^{(N)}}} = NP. \tag{4.32}$$

Hence we can express δ_1 as

$$\sqrt{\delta_1} = \frac{\sum_{i=1}^{N_t} \sqrt{\lambda_i^{(N)}}}{NP}. \tag{4.33}$$

Finally, plug (4.33) back into (4.31), we have the solution for the relaxed optimization problem given by

$$\mu_i = \frac{NP}{\sum_{j=1}^{N_t} \sqrt{\lambda_j^{(N)}}} \cdot \frac{1}{\sqrt{\lambda_i^{(N)}}}. \tag{4.34}$$

CHAPTER 5
FEEDBACK DESIGN

In order to obtain the advantage of employing the optimal training sequence set, information about the noise has to be measured at the receiver and fed back to the transmitter so that it can construct the optimal sequence set. The obvious questions are what information about the jammer we should feed back to the transmitter, and whether this feedback design is practical or not. We study these questions in this chapter.

5.1 Feedback Design for BLUE

From the Theorem 7 in Chapter 4, we see that the optimal training sequence set depends only on the eigen-structure of the time correlation matrix, $\mathbf{Q}_N$, of the jammer when we employ the BLUE as our channel estimator. Since $\mathbf{Q}_N$ varies much slower than the channel matrix, it is possible to estimate relevant information about $\mathbf{Q}_N$ at the receiver and feed back this information to the transmitter for it to construct and use the optimal training sequence set. It is desirable that if we can reduce the estimation error of the short-term channel fading parameters by using a minimal amount of information that is fed back from the receiver. In this section, we develop such a feedback scheme based on the fact that a suitable Toeplitz matrix can be approximated by a circulant matrix. In addition, the asymptotic behavior of the matrices, discussed earlier in Chapter 2, is considered to design feedback scheme.

When the jammer process is wide-sense stationary, time correlation matrix $\mathbf{Q}_N$ with BLUE approach take the form of a Toeplitz matrix. Indeed, consider a sequence of complex numbers $\{q_l\}_{l=-\infty}^{\infty}$ such that $q_l = q_{-l}^*$. The elements of $\mathbf{Q}_N$ at the ith row and jth column is given by q_{i-j}. The sequence $\{q_l\}_{l=-\infty}^{\infty}$ is obtained by sampling the autocorrelation function of the jammer process at the symbol rate. In addition,

as we discussed in Section 2.4.3 in Chapter 2, if the sequence $\{q_l\}_{l=-\infty}^{\infty}$ is absolutely summable that the Toeplitz matrix $\mathbf{Q}_N$ can be approximated by the circulant matrix [27]

$$\hat{\mathbf{Q}}_N = \mathbf{F}_N \hat{\mathbf{\Lambda}}_N \mathbf{F}_N^H, \tag{5.1}$$

where $\mathbf{F}_N$ is the $N \times N$ discrete Fourier transform matrix, i.e., the (k,l)th element of $\mathbf{F}_N$ is $\frac{1}{\sqrt{N}} e^{-j\frac{2\pi(k-1)(l-1)}{N}}$. $\mathbf{\Lambda}$ is $N \times N$ diagonal matrix with $\hat{\delta}_1^{(N)}, \ldots, \hat{\delta}_N^{(N)}$ as its diagonal elements.

A reasonable ways, as discussed in Section 2.4.3, to obtain $\delta_l^{(N)}$, for $l = 1, \ldots, N$, is

$$\delta_l^{(N)} = \sum_{k=1}^{N} (q_{1-k} + q_{N-k+1}) e^{-j\frac{2\pi(k-1)(l-1)}{N}}. \tag{5.2}$$

With this choice of $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$, it can be shown in Section 2.4.3 that $\hat{\mathbf{Q}}_N$ approaches $\mathbf{Q}_N$ as N approaches infinity. Moreover, if we arrange $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$ in an ascending order, we have

$$\lim_{N\to\infty} \left| \delta_{[l]}^{(N)} - \lambda_l^{(N)} \right| = 0, \tag{5.3}$$

for $l = 1, \ldots, n_t$. $\delta_{[l]}^{(N)}$ is the lth smallest eigenvalue among the set $\{\delta_l^{(N)}\}_{l=1}^{N}$. In (5.3), $\lambda_1^{(N)}, \cdots, \lambda_{n_t}^{(N)}$, defined in Chapter 4, are the n_t smallest eigenvalues of $\mathbf{Q}_N$. In summary, if we can estimate the autocorrelation function of the interference at the receiver, we can obtain $\delta_l^{(N)}$ by (5.2). Then the n_t smallest $\delta_l^{(N)}$'s and the corresponding indices are fed back to the transmitter. They correspond to the n_t frequency components of the noise that have the smallest power. At the transmitter, we can replace $\lambda_1^{(N)}, \ldots, \lambda_{n_t}^{(N)}$ and $\bar{\mathbf{U}}_N$, defined in Chapter 4, by the n_t frequency components that have the smallest power. As a result, the receiver only need to feedback the power values and indices of these n_t frequency components to the transmitter for the construction of the approximate optimal training sequence set.

5.2 Feedback Design for Bayesian Estimator

In order to obtain the advantage of the optimal training sequence design, long-term statistics of the interference correlation need to be estimated at the receiver and fed back to the transmitter. We can conclude from Section 4.2 that the optimal training sequence set depends on the channel gain variance σ^2 and the eigen-structure of the matrix $\mathbf{A}_N$ defined in (3.24) in Chapter 3. As a result, only these two long-term statistics need to be estimated at the receiver. Since the interferer signals are wide-sense stationary, $\mathbf{A}_N$ takes the form of a Toeplitz matrix. Indeed, consider a sequence of complex numbers $\{a_l\}_{l=-\infty}^{\infty}$ such that $a_l = a_{-l}^*$ and the elements of $\mathbf{A}_N$ at the ith row and jth column is given by a_{i-j}. The sequence $\{a_l\}_{l=-\infty}^{\infty}$ is obtained by sampling the autocorrelation function of the interference at the symbol rate. In addition, as we discussed in Section 2.4.3, if the sequence $\{a_l\}_{l=-\infty}^{\infty}$ is absolutely summable, then it is shown in Gray [27] that the Toeplitz matrix $\mathbf{A}_N$ can be approximated by the circulant matrix

$$\tilde{\mathbf{A}}_N = \mathbf{F}_N \mathbf{\Lambda}_N \mathbf{F}_N^H, \tag{5.4}$$

where $\mathbf{F}_N$ is the $N \times N$ FFT matrix, i.e., the (k,l)th element of $\mathbf{F}_N$ is $\frac{1}{\sqrt{N}} e^{-j\frac{2\pi(k-1)(l-1)}{N}}$, and $\mathbf{\Lambda}$ is a $N \times N$ diagonal matrix with $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$ as its diagonal elements.

To obtain $\delta_l^{(N)}$, for $l = 1, \ldots, N$, we use the same way in (5.2). With this choice of $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$, it can be shown that $\tilde{\mathbf{A}}_N$ approach $\mathbf{A}_N$ as N approach infinity [27]. Moreover if we arrange $\delta_1^{(N)}, \ldots, \delta_N^{(N)}$ in an ascending order, we have

$$\lim_{N \to \infty} |\delta_{[l]}^{(N)} - \lambda_l^{(N)}| = 0, \tag{5.5}$$

for $l = 1, \ldots, n_t$, and $\delta_{[l]}^{(N)}$ is the lth smallest eigenvalue among the set $\{\delta_l^{(N)}\}_{l=1}^{N}$.

Now we turn to the estimation of σ^2 and $\{a_k\}_{k=0}^{N-1}$. As mentioned before, they are both long-term statistics and hence should be estimated based on the observed training frames of the previous K packets, where K is smaller than the number of packets during which the long-term statistics remain the same. Toward this end,

let $s_{i,l}(n)$ denote the (i,l)th element of the training matrix $\mathbf{S}(n)$ and $y_{i,l}(n)$ denote the (i,l)th element of the observed training matrix $\mathbf{Y}(n)$ as defined in (3.14) for the nth packet, respectively. Then it is not hard to see that for $i = 1, 2, \ldots, n_r$ and $k = 0, 1, \ldots, N-1$,

$$(N-k)a_k = E\left[\frac{1}{K}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k} y_{i,l}(j)y_{i,l+k}^*(j) - \sigma^2 N R_n^s(k)\right], \tag{5.6}$$

where

$$R_n^s(k) = \frac{1}{NK}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k}\sum_{m=1}^{n_t} s_{m,l}(j)s_{m,l+k}^*(j). \tag{5.7}$$

In addition, let $\mathbf{P_S}(n)$ be the $N\times N$ projection matrix onto the subspace perpendicular to the one spanned by the rows of the training matrix $\mathbf{S}(n)$ for the nth packet. Denote the (i,l)th element of $\mathbf{Y}(n)\mathbf{P_S}(n)$ by $\tilde{y}_{i,l}(n)$ and the (k,l)th element of $\mathbf{P_S}(n)$ by $p_{k,l}(n)$. Then it can be shown that for $i = 1, 2, \ldots, n_r$,

$$a_0 R_n^p(0) + 2\sum_{k=1}^{N-1}\mathrm{Re}[a_k R_n^p(k)] = E\left[\frac{1}{K}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N}|\tilde{y}_{i,l}(j)|^2\right], \tag{5.8}$$

where

$$R_n^p(k) = \frac{1}{K}\sum_{j=n-K+1}^{n}\sum_{l=1}^{N-k} p_{l,l+k}(j). \tag{5.9}$$

We note that (5.6) for $k = 0, \ldots, N-1$ together with (5.8) provide us $2N$ equations to solve for the $2N$ unknowns σ^2, a_0 (both real-valued), and a_k for $k = 1, \ldots, N-1$ (all complex-valued). Thus estimates of σ^2 and a_k for $k = 0, 1, \ldots, N-1$ can be obtained by solving this set of linear equations with the expectation terms

replaced by their usual estimates as follows:

$$
\begin{aligned}
a_0 + \sigma^2 R_n^s(0) &= \frac{1}{n_r NK} \sum_{i=1}^{n_r} \sum_{j=n-K+1}^{n} \sum_{l=1}^{N} |y_{i,l}(j)|^2 \\
&\vdots \\
a_k + \sigma^2 R_n^s(k) &= \frac{1}{n_r NK} \sum_{i=1}^{n_r} \sum_{j=n-K+1}^{n} \sum_{l=1}^{N-k} y_{i,l}(j) y_{i,l+k}^*(j) \\
&\vdots \\
a_{N-1} + \sigma^2 R_n^s(N-1) &= \frac{1}{n_r NK} \sum_{i=1}^{n_r} \sum_{j=n-K+1}^{n} y_{i,1}(j) y_{i,N}^*(j) \\
a_0 R_n^p(0) + 2 \sum_{k=1}^{N-1} \mathrm{Re}[a_k R_n^p(k)] &= \frac{1}{n_r K} \sum_{i=1}^{n_r} \sum_{j=n-K+1}^{n} \sum_{l=1}^{N} |\bar{y}_{i,l}(j)|^2. \qquad (5.10)
\end{aligned}
$$

In above, the biased estimators

$$
\frac{N-k}{N} \left[\frac{1}{n_r} \sum_{i=1}^{n_r} \sum_{j=n-K+1}^{n} \sum_{l=1}^{N-k} y_{i,l}(j) y_{i,l+k}^*(j) - \sigma^2 R_n^s(k) \right],
$$

for $k = 0, 1, \ldots, N-1$, have been employed to approximate the corresponding expectations on the right hand side of (5.6). We note that the use of these biased estimators is similar to the use of biased autocorrelation function estimators in the Yule-Walker method of estimating the power spectral density of an AR process [30].

In summary, we can use the solution of (5.10) to estimate the autocorrelation function of the interference at the receiver and obtain estimates of the $\delta_l^{(N)}$'s by (5.2). Then the estimated value of σ^2, the n_t smallest $\delta_l^{(N)}$'s, and the corresponding indices are fed back to the transmitter. At the transmitter, we can replace $\bar{\mathbf{U}}_N$ by the columns of $\mathbf{F}_N$ that are indexed by the feedback indices to construct the optimal training sequence set for the current packet.

We note that the estimates of the a_k's obtained by solving (5.10) do not guarantee the resulting estimates of the δ_k's and σ^2 to be positive, although this is almost always the case when N, K, and n_r are sufficiently large. When the estimate of σ^2 is negative, we heuristically use the absolute value of the estimate instead. In addition,

we do not use those δ_k's with negative estimates in finding the n_t minimum values of $\lambda_1^{(N)}, \ldots, \lambda_N^{(N)}$ as described before.

5.3 Summary

In this chapter, we studied the way to feed back the required information from the receiver so that we can obtain approximate optimal training sequence sets. This feedback scheme is constitued based on the fact that a suitable Toeplitz matrix can be approximated by a circulant matrix. In the following chapter, the performance of the approximate optimal training sequence set obtained by the proposed feedback scheme will be compared with different training sequence sets in terms of the MSE of the channel estimator.

CHAPTER 6
NUMERICAL RESULTS

In this chapter, we present numerical results for the MSEs of the BLUE and Bayesian channel estimator when three different training sequence sets are employed. In addition, we introduce the asymptotic maximum MSE reduction ratio that can tell us the maximum advantage we can obtain by employing the optimal training sequence set over other choices of training sequences. Finally, conclusions will be given based on these numerical results.

6.1 Asymptotic Estimation Performance Gain

It is illustrative as well as practical to develop a simple measure that can tell us how much advantage we can obtain by employing the optimal training sequence set over other choices of training sequences. For instance, if the receiver determines that there is no much to be gain by using the optimal training sequences, it can inform the transmitter to keep on using the current ones. To this end, we employ equal-power orthogonal training sequences as our baseline for comparison, since these training sequences are commonly [18-21] suggested when the noise is white.

6.1.1 BLUE

First, we want to obtain the worst-case MSE, $\mathrm{MSE}_{\max}^{(N)}$, when equal-power orthogonal training sequences are employed and the total transmit power is as least P. It is not too hard to see that

$$\begin{aligned} \max_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \operatorname{tr}(\mathbf{S}\mathbf{Q}_N^{-1}\mathbf{S}^H)^{-1} &\leq \frac{n_t}{\min\limits_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \lambda_{\min}(\mathbf{S}\mathbf{Q}_N^{-1}\mathbf{S}^H)} = \frac{n_t^2\lambda_N^{(N)}}{NP} \\ \max_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \operatorname{tr}(\mathbf{S}\mathbf{Q}_N^{-1}\mathbf{S}^H)^{-1} &\geq \frac{n_t}{\min\limits_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \lambda_{\max}(\mathbf{S}\mathbf{Q}_N^{-1}\mathbf{S}^H)} = \frac{n_t^2\lambda_{N-n_t+1}^{(N)}}{NP}, \end{aligned} \tag{6.1}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are the minimum and maximum eigenvalues of a Hermitian matrix, respectively, and $\lambda_1^{(1)}, \ldots, \lambda_N^{(N)}$ are the eigenvalues of $\mathbf{Q}_N$ arranged in an ascending order. From Theorem 5, we can bound the worst-case MSE as

$$\frac{n_t^2 \lambda_{N-n_t+1}^{(N)}}{NP} \leq \frac{\mathrm{MSE}_{\max}^{(N)}}{\mathrm{tr}(\mathbf{Q}_r)} \leq \frac{n_t^2 \lambda_N^{(N)}}{NP}. \tag{6.2}$$

On the other hand, Theorem 7 shows that when the optimal training sequence set is employed,

$$\frac{n_t^2 \lambda_1^{(N)}}{NP} \leq \frac{\mathrm{MSE}_{*}^{(N)}}{\mathrm{tr}(\mathbf{Q}_r)} = \frac{\left(\sum_{i=1}^{n_t} \sqrt{\lambda_i^{(N)}}\right)^2}{NP} \leq \frac{n_t^2 \lambda_{n_t}^{(N)}}{NP}. \tag{6.3}$$

Combining (6.8) and (6.9), we can bound the ratio between the minimum MSE and the worst-case MSE by

$$\frac{\lambda_1^{(N)}}{\lambda_N^{(N)}} \leq \frac{\mathrm{MSE}_{*}^{(N)}}{\mathrm{MSE}_{\max}^{(N)}} \leq \frac{\lambda_{n_t}^{(N)}}{\lambda_{N-n_t+1}^{(N)}}. \tag{6.4}$$

This MSE ratio gives the maximum possible relative reduction in the estimation error that we can obtain by using the optimal sequence set under a specific jammer. To obtain a simpler performance metric, we consider the asymptotic value of this MSE ratio when N is very large.

Theorem 8. *Suppose that the sampled autocorrelation sequence, $\{q_n\}_{n=-\infty}^{\infty}$, of the wide-sense stationary jammer process is absolutely summable. Let*

$$Q(\omega) = \sum_{n=-\infty}^{\infty} q_n e^{-jn\omega}$$

be the discrete-time Fourier transform of $\{q_n\}_{n=-\infty}^{\infty}$. Then, for $i = 1, 2, \ldots, n_t$,

$$\begin{aligned} \lim_{N\to\infty} \lambda_i^{(N)} &= \min_{0\leq\omega\leq 2\pi} Q(\omega) \\ \lim_{N\to\infty} \lambda_{N-i+1}^{(N)} &= \max_{0\leq\omega\leq 2\pi} Q(\omega). \end{aligned} \tag{6.5}$$

Moreover the asymptotic maximum MSE reduction ratio

$$\Gamma = \lim_{N\to\infty} \frac{MSE_{*}^{(N)}}{MSE_{\max}^{(N)}} = \frac{\min_{0\leq\omega\leq 2\pi} Q(\omega)}{\max_{0\leq\omega\leq 2\pi} Q(\omega)}. \tag{6.6}$$

Proof. The results in (6.11) regarding the extremal eigenvalues of the sequence of Toeplitz matrices $\{\mathbf{Q}_N\}$ are proved in Grenander [36, Ch.5]. Applying (6.5) to (6.4), we get (6.6). □

6.1.2 Bayesian Estimatior

Similarly, we want to obtain the worst-case MSE, $\mathrm{MSE}_{\max}^{(N)}$, when equal-power orthogonal training sequences are employed and the total transmit power is as least P. It is not too hard to see that

$$\begin{aligned}
\max_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \operatorname{tr}\left[\mathbf{S}\mathbf{A}_N^{-1}\mathbf{S}^H+\frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1} &\leq \frac{n_t}{\min\limits_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \lambda_{\min}(\mathbf{S}\mathbf{A}_N^{-1}\mathbf{S}^H)+\frac{1}{\sigma_H^2}} \\
&= \frac{n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_N^{(N)}}+\frac{1}{\sigma^2}} \\
\max_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \operatorname{tr}\left[\mathbf{S}\mathbf{A}_N^{-1}\mathbf{S}^H+\frac{1}{\sigma^2}\mathbf{I}_{n_t}\right]^{-1} &\geq \frac{n_t}{\min\limits_{\mathbf{S}\mathbf{S}^H=NP\mathbf{I}} \lambda_{\max}(\mathbf{S}\mathbf{A}_N^{-1}\mathbf{S}^H)+\frac{1}{\sigma^2}} \\
&= \frac{n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_{N-n_t+1}^{(N)}}+\frac{1}{\sigma^2}},
\end{aligned} \tag{6.7}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are the minimum and maximum eigenvalues of a Hermitian matrix, respectively, and $\lambda_1^{(1)},\ldots,\lambda_N^{(N)}$ are the eigenvalues of $\mathbf{A}_N$ arranged in an ascending order. From (3.26), we can bound the worst-case MSE as

$$\frac{n_r n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_{N-n_t+1}^{(N)}}+\frac{1}{\sigma^2}} \leq \mathrm{MSE}_{\max}^{(N)} \leq \frac{n_r n_t}{\frac{NP}{n_t}\cdot\frac{1}{\lambda_N^{(N)}}+\frac{1}{\sigma^2}}. \tag{6.8}$$

On the other hand, from (4.15), when the optimal training sequence set is employed, the minimum MSE can be bounded by

$$\frac{n_r n_*}{\frac{NP}{n_*}\frac{1}{\lambda_1^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_{n_*}^{(N)}}{\lambda_1^{(N)}}} \leq \mathrm{MSE}_*^{(N)} \leq \frac{n_r n_*}{\frac{NP}{n_*}\frac{1}{\lambda_{n_*}^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_{n_*}^{(N)}}{\lambda_{n_*}^{(N)}}}. \tag{6.9}$$

Combining (6.8) and (6.9), we can bound the ratio between the minimum MSE and the worst-case MSE by

$$\frac{n_*}{n_t}\cdot\frac{\frac{1}{\lambda_N^{(N)}}+\frac{1}{\sigma^2}\frac{n_t}{NP}}{\frac{1}{\lambda_1^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_{n_*}^{(N)}}{\lambda_1^{(N)}}\frac{n_t}{NP}}\leq\frac{\mathrm{MSE}_*^{(N)}}{\mathrm{MSE}_{\max}^{(N)}}\leq\frac{n_*}{n_t}\cdot\frac{\frac{1}{\lambda_{N-n_t+1}^{(N)}}+\frac{1}{\sigma^2}\frac{n_t}{NP}}{\frac{1}{\lambda_{n_*}^{(N)}}+\frac{1}{\sigma^2}\frac{\lambda_1^{(N)}}{\lambda_{n_*}^{(N)}}\frac{n_t}{NP}}. \tag{6.10}$$

This MSE ratio gives the maximum possible relative reduction in the estimation error that we can obtain by using the optimal sequence set under a specific set of interferers. Here we obtain a simpler performance metric by considering the asymptotic value of this MSE ratio when N is very large.

As described in the Theorem 8, we employ the following results regarding the extremal eigenvalues of the sequence of Toeplitz matrices $\{\mathbf{A}_N\}$. Suppose that the sampled autocorrelation sequence, $\{a_n\}_{n=-\infty}^{\infty}$, of the wide-sense stationary interference process is absolutely summable. Let

$$A(\omega)=\sum_{n=-\infty}^{\infty}a_n e^{-jn\omega}$$

be the discrete-time Fourier transform of $\{a_n\}_{n=-\infty}^{\infty}$. Then, for $i=1,2,\ldots,n_t$,

$$\begin{aligned}\lim_{N\to\infty}\lambda_i^{(N)} &= \min_{0\leq\omega\leq 2\pi}A(\omega)\\ \lim_{N\to\infty}\lambda_{N-i+1}^{(N)} &= \max_{0\leq\omega\leq 2\pi}A(\omega).\end{aligned} \tag{6.11}$$

From (4.13) and (6.11), we see that $\lim_{N\to\infty} n_* = n_t$. Applying this and (6.11) to (6.10), the asymptotic maximum MSE reduction ratio

$$\Gamma=\lim_{N\to\infty}\frac{\mathrm{MSE}_*^{(N)}}{\mathrm{MSE}_{\max}^{(N)}}=\frac{\min\limits_{0\leq\omega\leq 2\pi}A(\omega)}{\max\limits_{0\leq\omega\leq 2\pi}A(\omega)}. \tag{6.12}$$

6.2 Numerical Examples for BLUE

In this section, we introduce two jammer models to illustrate the potential advantage of employing the optimal training sequence set. The first model considers the case when the jammer signal is a first-order auto-regressive (AR) random process. The

second model considers the case where the jammer is a co-channel interferer whose signal structure is exactly the same as that of the desired signal. In each model, we evaluate the MSEs of the channel estimators when following three different training sequence sets are employed:

1. Hadamard sequence set, i.e., first two rows of a Hadamard matrix are used as the training sequences;
2. the optimal training sequence set described in Chapter 4. and
3. the approximate optimal training sequence set described in Chapter 5.

In each case, we assume that the desired user has two transmit antennas and two receive antennas. Moreover, we assume that there is a single interferer. In each example, we evaluate the MSEs, normalized by $\mathrm{tr}(\mathbf{Q}_r)$ (c.f. Eqn. (3.13)), of the BLUE channel estimator when the three different training sequence sets are employed. In each case, we consider different lengths of the training sequences (N = 16, 32, 64, 128, 256, 512, and 1024) and different signal-to-interference ratio (SIR = 0dB, -15dB, and -30dB).

6.2.1 AR(1) Jammer

We assume that the jammer is modeled by a first-order auto-regressive (AR) process with AR parameter α that can be interpreted as the intensity of correlation among the jammer symbols, i.e., with a larger value of α, the crosscorrelations among jammer symbols are larger. The AR model is given [37] by

$$s_t = \alpha s_{t-1} + u_t, \tag{6.13}$$

where u_t is white Gaussian noise with zero mean, and variance σ_u^2. It is easy to verify that for this jammer,

$$Q(\omega) = \frac{\sigma_u^2}{1 + \alpha^2 - 2\alpha\cos\omega}. \tag{6.14}$$

Hence, the asymptotic maximum MSE reduction ratio is given by

$$\Gamma = \frac{\min\limits_{0\leq\omega\leq 2\pi} Q(\omega)}{\max\limits_{0\leq\omega\leq 2\pi} Q(\omega)} = \frac{\frac{\sigma_n^2}{(1+\alpha)^2}}{\frac{\sigma_n^2}{(1-\alpha)^2}} = \left(\frac{1-\alpha}{1+\alpha}\right)^2. \tag{6.15}$$

The MSEs of the BLUE channel estimator with the three different training sequence sets are shown and compared in Figs. 6–1 and 6–2 for the cases of $\alpha = 0.7$ and 0.9, respectively. From these figures, we observe that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set. Obviously, this is a desirable result, because it indicates that we can obtain almost same performance as that obtained by using the optimal training sequence set by feeding back much less information to the transmitter. In general, we see that the optimal training sequence set significantly outperforms the Hadamard sequence set for the two different values of α. The advantage of using the optimal training sequence set increases as the correlation parameter α increases. From (6.15), the asymptotic maximum MSE reduction ratio $\Gamma = -15$dB and -25.5dB for the cases of $\alpha = 0.7$ and 0.9, respectively. From Figs. 6–1 and 6–2, we see that the MSE reduction ratios obtained by using the optimal sequence set against the Hadamard sequence set are -12dB and -22dB for the cases of $\alpha = 0.7$ and 0.9, respectively. This means that the Hadamard sequence set gives MSEs that are not much lower than the worst-case values.

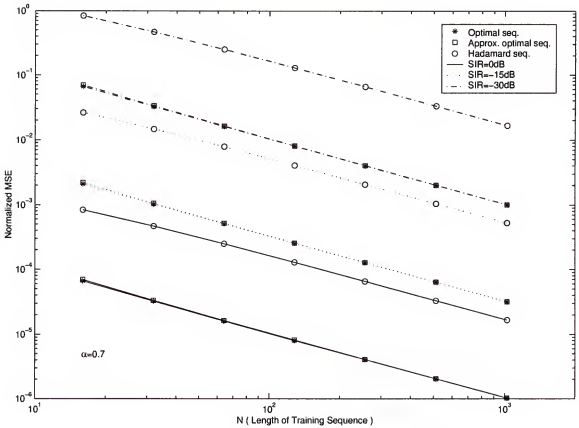


Figure 6–1: Comparison of MSEs obtained by using different training sequence sets. AR-1 jammer with $\alpha = 0.7$ and $n_t = n_r = 2$.

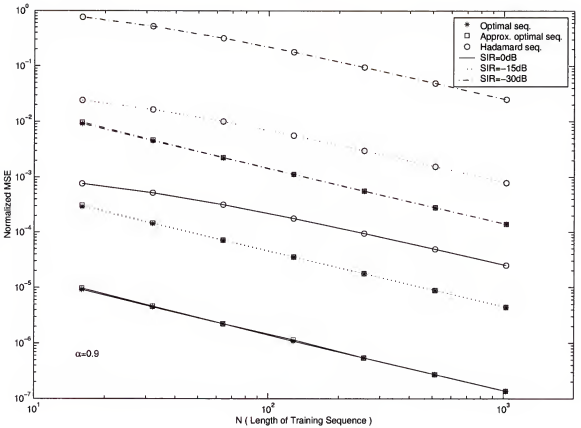


Figure 6–2: Comparison of MSEs obtained by using different training sequence sets. AR-1 jammer with $\alpha = 0.9$ and $n_t = n_r = 2$.

6.2.2 Co-channel Interferer

We assume that the jammer is a co-channel interferer whose signal format is exactly the same as the desired user signal. More precisely, let us assume that the transmitted signal at the kth transmit antenna of the jammer is given by

$$s_J^{(k)}(t) = \sqrt{P_J} \sum_{l=-\infty}^{\infty} b_l^{(k)} \psi(t - lT - \tau), \tag{6.16}$$

where $b_l^{(k)}$ is the sequence of data symbols that are assumed independent and identically distributed random variables with zero mean and unit variance, $\psi(t)$ is the symbol waveform, T is the symbol interval, and τ is the delay with respect to the timing of the desired signal. Without loss of generality, we can assume that $\tau \in [0, T)$. We also assume that $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$.

With the model described above, the elements of the jammer signal matrix $\mathbf{S}_J$ in (3.1) and (3.14) are samples at the matched filter output at the receiver at time iT. Specifically, the (k, i)th element of $\mathbf{S}_J$ is given by

$$[\mathbf{S}_J]_{k,i} = \sum_{l=-\infty}^{\infty} b_l^{(k)} \hat{\psi}((i - l)T - \tau), \tag{6.17}$$

where

$$\hat{\psi}(t) = \int_{-\infty}^{\infty} \psi(t - s)\psi^*(s) ds \tag{6.18}$$

is the autocorrelation of the symbol waveform. Thus, we can express sampled autocorrelation sequence as

$$\begin{aligned} q_n &= E\{[\mathbf{S}_J]_{k,m}[\mathbf{S}_J]_{k,m+n}\} \\ &= P_J \sum_{l=-\infty}^{\infty} \hat{\psi}((l - n)T - \tau)\hat{\psi}^*(lT - \tau), \end{aligned} \tag{6.19}$$

and its discrete-time Fourier transform as

$$\begin{aligned} Q(\omega) &= P_J \left| \sum_{n=-\infty}^{\infty} \hat{\psi}(nT+\tau) e^{-j\omega n} \right|^2 \\ &= \frac{P_J}{T^2} \left| \sum_{n=-\infty}^{\infty} \hat{\Psi}\left(\frac{\omega - 2\pi k}{T}\right) e^{\frac{j(\omega-2\pi k)\tau}{T}} \right|^2, \end{aligned} \tag{6.20}$$

where $\hat{\Psi}(\Omega) = |\Psi(\Omega)|^2$, and $\Psi(\Omega)$ is the Fourier transform of the symbol waveform $\psi(t)$.

To illustrate how the use of the optimal training sequence set can benefit the channel estimation process, let us consider the following two common symbol waveforms:

6.2.2.1 Rectangular symbol waveform

In this case,

$$\psi(t) = \begin{cases} T, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{\psi}(t) = \begin{cases} t/T, & 0 \le t \le T \\ 2 - t/T, & T \le t \le 2T \\ 0, & \text{otherwise.} \end{cases}$$

From (6.20), we have

$$Q(\omega) = P_J[\frac{\tau^2}{T^2} + \frac{(T-\tau)^2}{T} + \frac{2\tau}{T}(\frac{T-\tau}{T})\cos\omega]. \tag{6.21}$$

Hence, the asymptotic maximum MSE reduction ratio is

$$\begin{aligned} \Gamma &= \frac{\min Q(\omega)}{\max Q(\omega)} \\ &= \left(\frac{2\tau - T}{T}\right)^2. \end{aligned} \tag{6.22}$$

From (6.22), the use of the optimal training sequence provides no gain when the co-channel interferer is symbol-synchronous to the desired user signal, i.e., $\tau = 0$. On the other hand, when $\tau = 0.5T$, the asymptotic maximum MSE reduction ratio is 0.

This means that we can almost completely eliminate the effect of the interferer by using the set of long optimal training sequences.

In this section, we compare the MSEs of the BLUE channel estimator with the three different training sequence sets in Figs. 6–3 and 6–4 for the cases of $\tau = 0.3T$ and $0.5T$, respectively. Again, we observe that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set. From (6.22), the asymptotic maximum MSE reduction ratio $\Gamma = -8$dB and $-\infty$dB for the cases of $\tau = 0.3T$ and $0.5T$, respectively. From Fig. 6–3, we see that the MSE reduction ratio obtained by using the optimal sequence set against the Hadamard sequence set is -4dB for the case of $\tau = 0.3T$. When $\tau = 0.5T$, the results in Fig. 6–4 indicate that the MSE reduction ratio decreases as N increases. These results are consistent with the asymptotic values predicted by (6.22).

6.2.2.2 ISI-free symbol waveform with raised cosine spectrum [38]

In this case,

$$\hat{\Psi}(\Omega) = \begin{cases} T, & 0 \le |\Omega| \le \frac{\pi(1-\beta)}{T} \\ \frac{T}{2}\left\{1 + \cos\left[\frac{T}{2\beta}\left(|\Omega| - \frac{\pi(1-\beta)}{T}\right)\right]\right\}, & \frac{\pi(1-\beta)}{T} \le |\Omega| \le \frac{\pi(1+\beta)}{T} \\ 0, & |\Omega| \ge \frac{\pi(1+\beta)}{T}, \end{cases}$$

where $0 < \beta \le 1$ is the roll-off factor. Since $\sum_{k=-\infty}^{\infty} \hat{\Psi}\left(\frac{\omega - 2\pi k}{T}\right) = T$ for all ω and $\hat{\Psi}(\Omega)$ is positive, it can be deduced from (6.20) that $\max_{0 \le \omega \le 2\pi} Q(\omega) = P_J$. To find $\min_{0 \le \omega \le 2\pi} Q(\omega)$, because of symmetry of $\hat{\Psi}(\Omega)$, it is enough to consider the interval $\omega \in [\pi(1-\beta), \pi]$. Over this interval, by (6.20), we have

$$Q(\omega) = \frac{P_J}{4}\left\{\left[1 + \cos\left[\frac{\omega - \pi(1-\beta)}{2\beta}\right]\right]^2 + \left[1 + \cos\left[\frac{\omega - \pi(1+\beta)}{2\beta}\right]\right]^2 + 2\cos\left(\frac{2\pi\tau}{T}\right)\left[1 + \cos\left[\frac{\omega - \pi(1-\beta)}{2\beta}\right]\right]\left[1 + \cos\left[\frac{\omega - \pi(1+\beta)}{2\beta}\right]\right]\right\}.$$

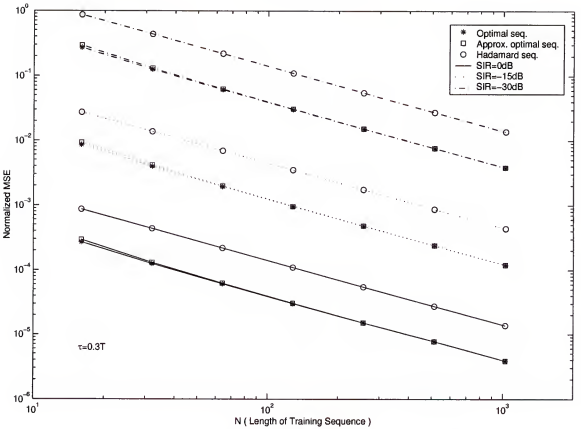


Figure 6–3: Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with rectangular waveform, $\tau = 0.3T$, and $n_t = n_r = 2$.

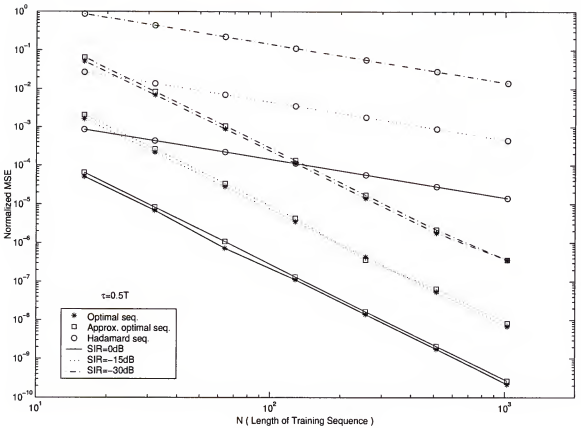


Figure 6–4: Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with rectangular waveform, $\tau = 0.5T$, and $n_t = n_r = 2$.

Simple calculus reveals that $\min_{0\leq\omega\leq 2\pi} Q(\omega) = \frac{P_J}{2}\left\{1+\cos\left(\frac{2\pi\tau}{T}\right)\right\}$. Thus, the asymptotic maximum MSE reduction ration is

$$\begin{aligned}\Gamma &= \frac{\min Q(\omega)}{\max Q(\omega)} \\ &= \frac{\frac{P_J}{2}\left\{1+\cos\left(\frac{2\pi\tau}{T}\right)\right\}}{P_J} \\ &= \cos^2\left(\frac{\pi\tau}{T}\right). \end{aligned} \tag{6.23}$$

From (6.23), the use of the optimal training sequence provides no gain when the co-channel interferer is symbol-synchronous to the desired user signal, i.e., $\tau = 0$. On the other hand, when $\tau = 0.5T$, the asymptotic maximum MSE reduction ratio is 0. This means that we can almost completely eliminate the effect of the interferer by using the set of long optimal training sequences.

In this section, we compare the MSEs of the BLUE channel estimator with the three different training sequence sets in Figs. 6–5 and 6–6 for the cases of $\tau = 0.3T$ and $0.5T$, respectively. Again, we observe that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set. From (6.23), the asymptotic maximum MSE reduction ratio $\Gamma = -4.6$dB and $-\infty$dB for the cases of $\tau = 0.3T$ and $0.5T$, respectively. From Fig. 6–5, we see that the MSE reduction ratio obtained by using the optimal sequence set against the Hadamard sequence set is -4dB for the case of $\tau = 0.3T$. When $\tau = 0.5T$, the results in Fig. 6–6 indicate that the MSE reduction ratio decreases as N increases. These results are consistent with the asymptotic values predicted by (6.23).

6.2.3 Average MSE with Hadamard Sequence Set

In this section, we evaluate the average MSE of the BLUE channel estimator obtained by using the different rows Hadamard matrix as training sequences. Then this result is compared with the MSE obtained by using optimal training sequence set. The MSEs with the Hadamard training sequence set in Figs. 6–1 - 6–6 were

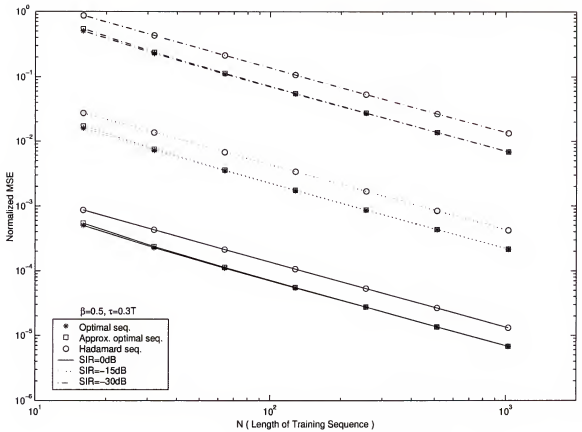


Figure 6–5: Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with raised-cosine waveform, $\beta = 0.5$, $\tau = 0.3T$, and $n_t = n_r = 2$.

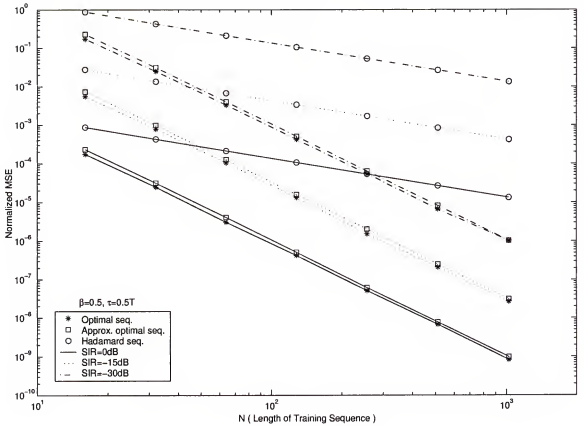


Figure 6–6: Comparison of MSEs obtained by using different training sequence sets. Co-channel interferer with raised-cosine waveform, $\beta = 0.5$, $\tau = 0.5T$, and $n_t = n_r = 2$.

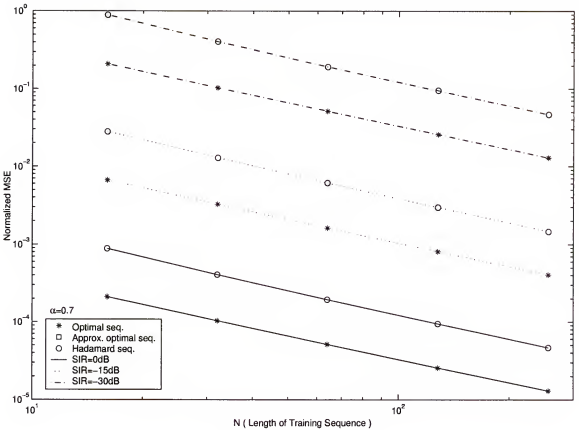


Figure 6–7: Comparison MSEs obtained by using optimal training sequence set with average MSEs obtained by using all possible Hadamard training sequence set. AR(1) jammer with $\alpha = 0.7$ and $n_t = n_r = 2$.

obtained by arbitrary choosing sequences from the first two rows of the Hadamard matrix. Because the MSE is dependent on the choice of the sequences, we average the MSEs obtained by using all possible Hadamard sequence pairs. In Figs. 6–7 and 6–8, we see that the optimal training sequence set still gives significantly smaller average MSE than using the rows of the Hadamard matrix as training sequences for the two different values of α. In Fig. 6–9, MSEs are compared when the rectangular waveform is considered as an interference signal. Like AR(1) model, we can still obtain significant performance gain by using optimal training sequence set over the Hadamard training sequence set.

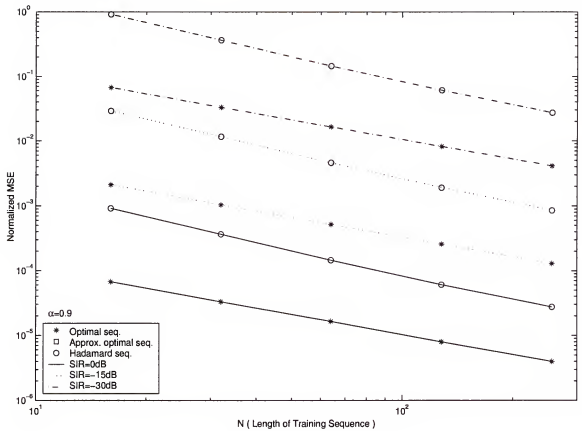


Figure 6–8: Comparison MSEs obtained by using optimal training sequence set with average MSEs obtained by using all possible Hadamard training sequence set. AR(1) jammer with $\alpha = 0.9$ and $n_t = n_r = 2$.

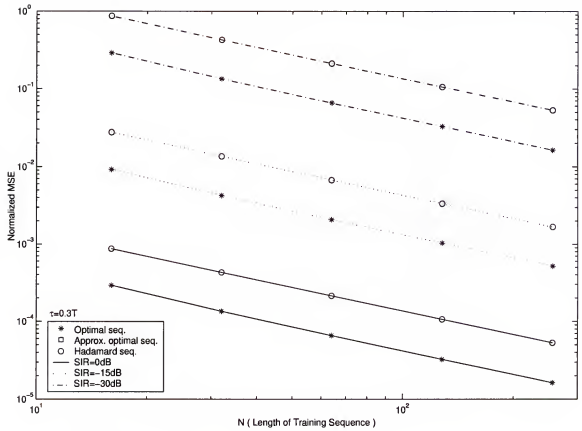


Figure 6–9: Comparison MSEs obtained by using optimal training sequence set with average MSEs obtained by using all possible Hadamard training sequence set. Co-channel interferer with rectangular waveform, $\tau = 0.3T$ and $n_t = n_r = 2$.

6.3 Numerical Examples for Bayesian Estimator

In this section, we evaluate the MSE, defined in (3.26), of the Bayesian channel estimator when the three different training sequence sets, described in previous section, are employed. We assume that there are two interferers in the system. We assume that each interferer transmits a different jamming signal generated with a different AR parameters or delay. We consider different lengths (N = 16, 32, 64, 128, 256, 512, and 1024) of the training sequences and different received signal-to-interference ratios ($\frac{\sigma_i^2 P_i}{\sigma_i^2 P_i}$ = 0dB and -20dB for i =1 and 2). The received signal-to-noise ratio, $\frac{\sigma^2 P}{\sigma_w^2}$ is set to 10dB. The proposed feedback design described in Section 5.2 is employed to obtain the approximate optimal training sequences. We have assumed that the received training signals from 10 previous packets are employed to estimate the jammer information at the receiver. The training sequences used in the previous 10 packets are the Hadamard sequences described in previous section.

6.3.1 AR(1) Jammer

We assume that there are two jammers in the system. Both jammers have one transmit antenna. The interference signals from the jammers are modeled by two first-order auto-regressive (AR) processes with AR parameters $0 < \alpha_1, \alpha_2 < 1$, respectively. For instance, the AR model of the first jammer is given by

$$s_t^{(1)} = \alpha s_{t-1}^{(1)} + u_{1,t}, \tag{6.24}$$

where $u_{1,t}$ is white Gaussian random process with zero mean and variance $\sigma_{u,1}^2$. The AR parameter can be interpreted as the intensity of correlation among the symbols of the jammer. It is easy to verify that for this case,

$$A(\omega) = \sum_{m=1}^{2} \frac{\sigma_m^2 \sigma_{u,m}^2}{1 + \alpha_m^2 - 2\alpha_m \cos\omega} + \sigma_w^2. \tag{6.25}$$

Hence, the asymptotic maximum MSE reduction ratio is given by

$$\Gamma = \frac{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \frac{1-\alpha_m}{1+\alpha_m} + 1}{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \frac{1+\alpha_m}{1-\alpha_m} + 1}, \tag{6.26}$$

where $P_m = \frac{\sigma_{u,m}^2}{1-\alpha_m^2}$ is the transmit power of the mth jammer.

From the Figs. 6–10 - 6–11, we see that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set. Obviously, this is a desirable result, because it indicates that we can obtain almost same performance as that obtained by using the optimal training sequence set by estimating the jammer information at the receiver and feeding back only a small amount of information to the transmitter. In general, we see that the optimal training sequence set significantly outperforms the Hadamard sequence set in all the cases considered. The advantage of using the optimal training sequence set increases as the correlation parameters α_i's increase. The asymptotic maximum MSE reduction ratios for the cases considered above are shown in Table 6–1. For comparison, the MSE reduction ratios obtained by using the optimal sequence set against the Hadamard sequence set for $N = 1024$ are also included in Table 6–1. We can deduce from the table that the Hadamard sequence set is rather inefficient. In addition, much more reduction in the MSE can be obtained using the optimal sequence set when both of the α_i's are close to 1.

6.3.2 Co-channel interferer

In this example, we assume that the interference is caused by two co-channel interferers whose signal format is similar to that of the desired user. More precisely, let us assume that the transmitted signal at the ith transmit antenna of the mth interferer is given by

$$s_i^{(m)}(t) = \sqrt{\frac{P_m}{2}} \sum_{l=-\infty}^{\infty} b_{i,l}^{(m)} \psi(t - lT - \tau_m), \tag{6.27}$$

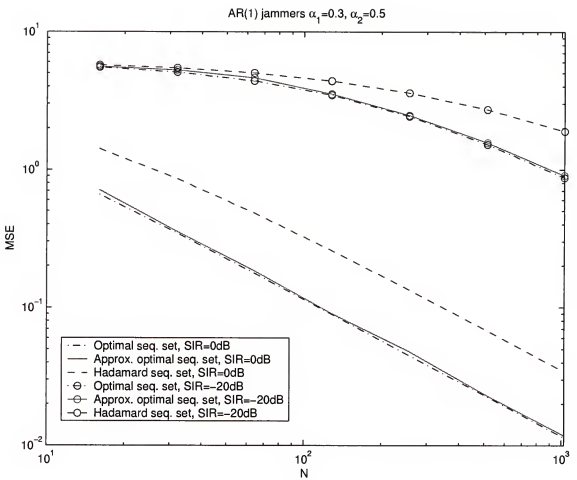


Figure 6–10: Comparison of MSEs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.3$ and $\alpha_2 = 0.5$.

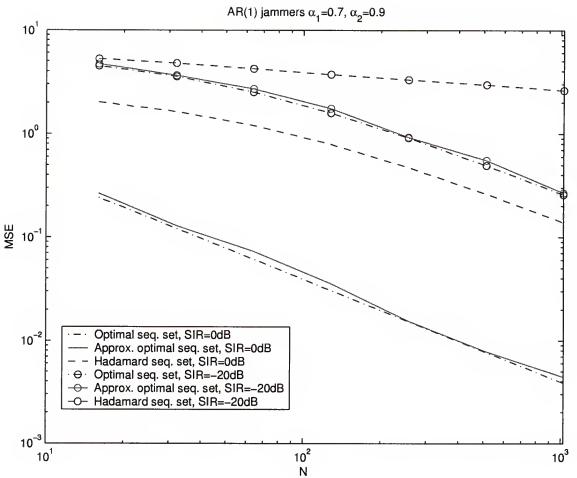


Figure 6–11: Comparison of MSEs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$.

where $b_{i,l}^{(m)}$ is the sequence of data symbols, which are assumed to be i.i.d. binary random variables with zero mean and unit variance, from the ith antenna of the mth interferer, $\psi(t)$ is the symbol waveform, T is the symbol interval, and τ_m is the symbol timing difference between the mth interferer and the desired signal. Without loss of generality, we can assume that $\tau_m \in [0, T)$. We also assume that $\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$.

With the model described above, the elements of the interference signal matrix $\mathbf{S}_m$ in (3.14) are samples at the matched filter output at the receiver at time kT. Specifically, the (i,k)th element of $\mathbf{S}_m$ is given by

$$s_{i,k}^{(m)} = \sum_{l=-\infty}^{\infty} b_{i,l}^{(m)} \hat{\psi}((k-l)T - \tau_m), \tag{6.28}$$

where

$$\hat{\psi}(t) = \int_{-\infty}^{\infty} \psi(t-s)\psi^*(s)ds \tag{6.29}$$

is the autocorrelation of the symbol waveform. Thus, it is easy to see that the sampled autocorrelation sequence

$$a_n = \sum_{m=1}^{2} \sigma_m^2 P_m \sum_{l=-\infty}^{\infty} \hat{\psi}((l-n)T - \tau_m)\hat{\psi}^*(lT - \tau) + \sigma_w^2 \delta_n, \tag{6.30}$$

and its discrete-time Fourier transform is given by

$$\begin{aligned} A(\omega) &= \sum_{m-1}^{2} \sigma_m^2 P_m \left| \sum_{n=-\infty}^{\infty} \hat{\psi}(nT + \tau_m) e^{-j\omega n} \right|^2 + \sigma_w^2 \\ &= \sum_{m-1}^{2} \frac{\sigma_m^2 P_m}{T^2} \left| \sum_{n=-\infty}^{\infty} \hat{\Psi}\left(\frac{\omega - 2\pi k}{T}\right) e^{\frac{j(\omega - 2\pi k)\tau_m}{T}} \right|^2 + \sigma_w^2, \end{aligned} \tag{6.31}$$

where P_m is the transmit power from the mth interferer, $\hat{\Psi}(\Omega) = |\Psi(\Omega)|^2$, and $\Psi(\Omega)$ is the Fourier transform of the symbol waveform $\psi(t)$.

To illustrate how the use of the optimal training sequence set can benefit the channel estimation process, let us consider the following two common symbol waveforms:

6.5.2.1 Rectangular symbol waveform

In this case,

$$\psi(t) = \begin{cases} 1/T, & 0 \le t \le T \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{\psi}(t) = \begin{cases} t/T, & 0 \le t \le T \\ 2 - t/T, & T \le t \le 2T \\ 0, & \text{otherwise.} \end{cases}$$

From (6.31), we have

$$A(\omega) = \sum_{m=1}^{2} \sigma_m^2 P_m \left[\left(\frac{\tau_m}{T} \right)^2 + \left(1 - \frac{\tau_m}{T} \right)^2 + 2 \left(\frac{\tau_m}{T} \right) \left(1 - \frac{\tau_m}{T} \right) \cos \omega \right] + \sigma_w^2. \quad (6.32)$$

Hence, the asymptotic maximum MSE reduction ratio is

$$\Gamma = \frac{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \left(2\frac{\tau_m}{T} - 1 \right)^2 + 1}{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} + 1}. \quad (6.33)$$

From (6.33), the use of the optimal training sequence provides no gain when the co-channel interferers are symbol-synchronous to the desired user signal, i.e., $\tau_1 = \tau_2 = 0$. On the other hand, when $\tau_1 = \tau_2 = 0.5T$, the asymptotic maximum MSE reduction ratio attains its minimum value. This means that we can almost completely eliminate the effect of the interferers by using the set of long optimal training sequences.

Like before, we compare the MSEs of the Bayesian channel estimator with the three different training sequence sets in Fig. 6–12 by considering the case in which $\tau_1 = 0.3T$ and $\tau_2 = 0.5T$. The other parameters are chosen as in the AR jammer example before. Again from Fig. 6–12, we observe that there exist only minimal differences between the MSEs of using the optimal training sequence set and approximate optimal training sequence set, and that the optimal training sequence set significantly outperforms the Hadamard sequence set in all the cases considered. The asymptotic maximum MSE reduction ratios for the cases considered above are shown

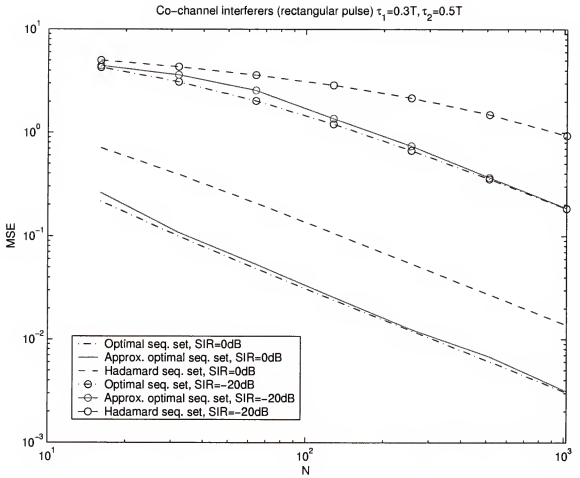


Figure 6–12: Comparison of MSEs obtained by using different training sequence sets. Two co-channel interferers with rectangular waveforms and delays $\tau_1 = 0.3T$, $\tau_2 = 0.5T$.

in Table 6–2. For comparison, the MSE reduction ratios obtained by using the optimal sequence set against the Hadamard sequence set for $N = 1024$ are also included in Table 6–2. We can deduce from the table that the Hadamard sequence set is rather inefficient.

6.5.2.2 ISI-free symbol waveform with raised cosine spectrum [38]

In this case,

$$\hat{\Psi}(\Omega) = \begin{cases} T, & 0 \le |\Omega| \le \frac{\pi(1-\beta)}{T} \\ \frac{T}{2}\left\{1 + \cos\left[\frac{T}{2\beta}\left(|\Omega| - \frac{\pi(1-\beta)}{T}\right)\right]\right\}, & \frac{\pi(1-\beta)}{T} \le |\Omega| \le \frac{\pi(1+\beta)}{T} \\ 0, & |\Omega| \ge \frac{\pi(1+\beta)}{T}, \end{cases}$$

where $0 < \beta \leq 1$ is the roll-off factor. Since $\sum_{k=-\infty}^{\infty} \hat{\Psi}\left(\frac{\omega - 2\pi k}{T}\right) = T$ for all ω and $\hat{\Psi}(\Omega)$ is positive, it can be deduced from (6.31) that $\max\limits_{0 \leq \omega \leq 2\pi} A(\omega) = \sum_{m=1}^{2} \sigma_m^2 P_m + \sigma_w^2$. To find $\min\limits_{0 \leq \omega \leq 2\pi} A(\omega)$, because of symmetry of $\hat{\Psi}(\Omega)$, it is enough to consider the interval $\omega \in [\pi(1-\beta), \pi]$. Over this interval, by (6.31), we have

$$A(\omega) = \sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{4} \left\{ \left[1 + \cos\left[\frac{\omega - \pi(1-\beta)}{2\beta} \right] \right]^2 + \left[1 + \cos\left[\frac{\omega - \pi(1+\beta)}{2\beta} \right] \right]^2 \right.$$
$$\left. + 2\cos\left(\frac{2\pi\tau_m}{T} \right) \left[1 + \cos\left[\frac{\omega - \pi(1-\beta)}{2\beta} \right] \right] \left[1 + \cos\left[\frac{\omega - \pi(1+\beta)}{2\beta} \right] \right] \right\} + \sigma_w^2. \tag{6.34}$$

Simple calculus reveals that $\min\limits_{0 \leq \omega \leq 2\pi} A(\omega) = \sum_{m=1}^{2} \sigma_m^2 P_m \cos^2\left(\frac{\pi\tau_m}{T}\right) + \sigma_w^2$. Thus, the asymptotic maximum MSE reduction ratio is

$$\Gamma = \frac{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} \cos^2\left(\frac{\pi\tau_m}{T}\right) + 1}{\sum_{m=1}^{2} \frac{\sigma_m^2 P_m}{\sigma_w^2} + 1}. \tag{6.35}$$

From (6.35), the use of the optimal training sequence provides no gain when the co-channel interferers are symbol-synchronous to the desired user signal, i.e., $\tau_1 = \tau_2 = 0$. On the other hand, when $\tau_1 = \tau_2 = 0.5T$, the asymptotic maximum MSE reduction ratio attains its minimum value. This means that we can almost completely eliminate the effect of the interferers by using the set of long optimal training sequences.

Like before, we compare the MSEs of the Bayesian channel estimator with the three different training sequence sets in Fig. 6–13 by considering the case in which $\tau_1 = 0.3T$ and $\tau_2 = 0.5T$. The roll-off factor of the ISI-free waveform is chosen to be $\beta = 0.5$. The other parameters are chosen as in the AR jammer example before. The conclusions from Fig. 6–13 are similar to those for the rectangular waveform.

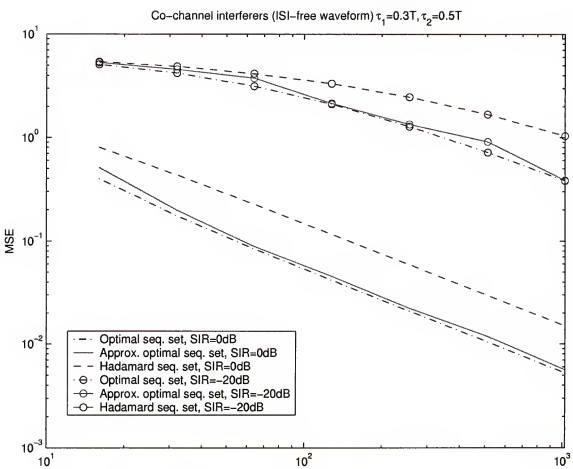


Figure 6–13: Comparison of MSEs obtained by using different training sequence sets. Two co-channel interferers with ISI-free waveform waveforms and delays $\tau_1 = 0.3T$, $\tau_2 = 0.5T$.

6.3.3 Bit Error Rate Performance

In this section, we evaluate the bit error rate (BER) performance by computer simulation when the Bayesian channel estimator and the three different training sequence sets are employed. We assume that there are two transmit antennas and four receive antennas. We also assume that there are two interferers in the system. Each having one transmit antenna. Interferers are modeled by two AR(1) processes with AR parameters α_1=0.7 and α_2=0.9, respectively. The training sequence length N is 64 and signal-to-interference ratios are 0dB and 10dB. We consider different values of signal-to-noise ratios from 0dB to 40dB. In this simulation, we employ Alamouti's space-time block code [8]. The goal of this simulation is to examine the relationship between the channel estimation performance obtained by using different training sequence sets and the corresponding BER performance.

Figs. 6–14 and 6–15 are the simulation results of the BER performance obtained by using different training sequence sets. From these figures, we observe that we can reduce BER significantly by using optimal training sequence sets against the Hadamard training sequence set. Moreover, we see that there exist only minimal differences between the BERs of using the optimal training sequence set and approximate optimal training sequence set obtained by proposed feedback scheme. This indicates that we can obtain almost same BER performance as that obtained by using the optimal training sequence set by feeding back much less information to the transmitter.

6.4 Conclusion

Numerical results showed that the MSEs of the BLUE and Bayesian channel estimator can be reduced significantly by using the optimized training sequence set over the Hadamard training sequence set that is usually used. In addition, we observed that we can obtain almost same performance by using the approximate optimal training sequence set obtained by the proposed feedback scheme.

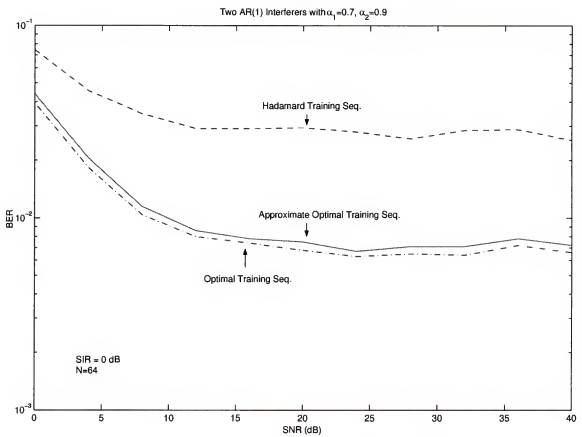


Figure 6–14: Comparison of BERs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$.

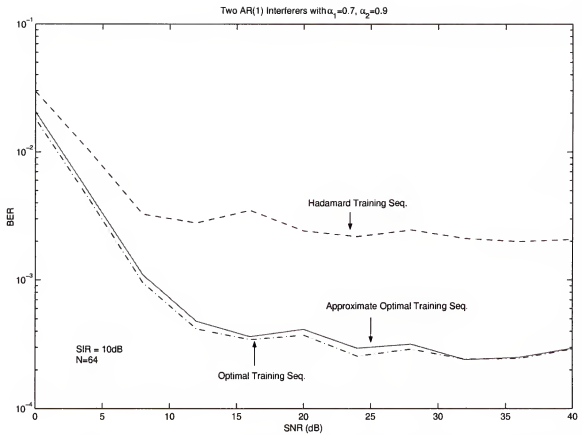


Figure 6–15: Comparison of BERs obtained by using different training sequence sets. Two AR(1) interferers with $\alpha_1 = 0.7$ and $\alpha_2 = 0.9$.

Table 6–1: Comparison of asymptotic maximum MSE reduction ratio and MSE ratio between using optimal and Hadamard sequences in the case of AR jammers.

$\alpha_1 = 0.3$, $\alpha_2 = 0.5$		
SIR	Asymp. max. MSE reduct. ratio	MSE with optimal seqs. / MSE with Hadamard seqs. $(N = 1024)$
0dB	−7.08dB	−4.81dB
−20dB	−7.46dB	−3.37dB

$\alpha_1 = 0.7$, $\alpha_2 = 0.9$		
SIR	Asymp. max. MSE reduct. ratio	MSE with optimal seqs. / MSE with Hadamard seqs. $(N = 1024)$
0dB	−18.77dB	−15.56dB
−20dB	−20.30dB	−10.05dB

Table 6–2: Comparison of asymptotic maximum MSE reduction ratio and MSE ratio between using optimal and Hadamard sequences in the case of co-channel interferers.

Rectangular waveform $\tau_1 = 0.3T$, $\tau_2 = 0.5T$		
SIR	Asymp. max. MSE reduct. ratio	MSE with optimal seqs. / MSE with Hadamard seqs. $(N = 1024)$
0dB	−9.07dB	−6.55dB
−20dB	−10.94dB	−7.08dB

ISI-free waveform $\tau_1 = 0.3T$, $\tau_2 = 0.5T$		
SIR	Asymp. max. MSE reduct. ratio	MSE with optimal seqs. / MSE with Hadamard seqs. $(N = 1024)$
0dB	−6.73dB	−4.54dB
−20dB	−7.62dB	−4.34dB

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this work, we addressed the problems of channel estimation and optimal training sequence design for multiple-input and multiple-output (MIMO) systems over flat fading channels in the presence of colored interference. We considered two different system models for which we applied two different channel estimators. The BLUE channel estimator was considered for the case where there is single interferer with the deterministic channel assumption. The Bayesian channel estimator was considered for the case where there are multiple interferers. In the Bayesian approach, known channel parameters that need to be estimated were assumed to be random with prior PDF. This consititues the major difference between the two cases considered. We showed that the MSEs of the considered channel estimators depend on the choice of the training sequence set. Based on this observation, we addressed the problem of optimal training sequence set design and obtained the optimal training sequence set under a total transmission power constraint. In order to obtain the advantage of the optimal training sequence design, we developed an information feedback scheme that required a minimal amount of information from the receiver to approximately construct the optimal training sequence set. Moreover, to obtain a simpler performance metric, we introduced the asymptotic maximum MSE reduction ratio that could tell us the maximum possible relative reduction in the estimation error that we could obtain by using the optimal training sequence set under specific jammer models.
Numerical results showed that the MSEs of the channel estimators for both BLUE and Bayesian estimator can be reduced significantly by using the optimal training sequence sets, proposed in Chapter 4, over the Hadamard training sequence sets. In

addition, we verified from the numerical results that the approximate optimal training sequence sets, obtained by the proposed feed back scheme in Chapter 5, gave estimation performance comparable to the optimal training sequence sets by utilizing minimal information from the receiver.

7.2 Future Work

As discussed in Chapter 1, there have been two major approaches to design optimal training sequence. One approach [9-13], [18], [21] is to determine the training sequence that minimizes the MSE of the channel estimator. The other approach [14], [20], [23] is to use information theoretic approach that uses lower bound of the channel capacity. Frequency-selective fading channels have been considered in Vikalo et al. [14], Fragouli et al. [21], Ma et al. [23] and all of these works use the information theoretic approach to obtain optimal training sequences that can maximize some lower bounds on the channel capacity. Most of these works have considered the presence of only white noise. Not much work has been done to design optimal training sequence in the presence of colored interference over frequency-selective fading channels.

In this dissertation, we have discussed the channel estimation and optimal training sequence design problems in the presence of colored noise over frequency flat fading channels. Therefore, it is natural to extend this work to design optimal training sequence in the presence of colored interferece over frequency-selective fading channels in MIMO systems.

Problem Approach

Let us consider the Bayesian channel estimator approach for the single interferer case for simplicity. Multiple interferer case can be easily extended from the single interferer case. The system model is given by

$$\mathbf{Y} = \mathbf{SH} + \underbrace{\mathbf{S}_J\mathbf{H}_J + \mathbf{W}}_{\mathbf{E}} = \mathbf{SH} + \mathbf{E}, \tag{7.1}$$

where the $N \times n_t L$ training symbol matrix $\mathbf{S}$ is defined as

$$\mathbf{S} = [\mathbf{S}_1 \ \mathbf{S}_2 \ \cdots \ \mathbf{S}_{n_t}], \tag{7.2}$$

and L is the delay spread of the channel in symbol period, the $N \times L$ matrix $\mathbf{S}_i$ is given by

$$\mathbf{S}_i = \begin{bmatrix} s_i(0) & 0 & \cdots & 0 \\ s_i(1) & s_i(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_i(L-1) & s_i(L-2) & \cdots & s_i(0) \\ \vdots & \vdots & \vdots & \vdots \\ s_i(N-1) & s_i(N-2) & \cdots & s_i(N-L) \end{bmatrix}, \tag{7.3}$$

for $i = 1, 2, \ldots, n_t$, and the $n_t L \times n_r$ channel matrix $\mathbf{H}$ is given as

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_r} \\ \vdots & & \vdots \\ h_{n_t,1} & \cdots & h_{n_t,n_r} \end{bmatrix} \tag{7.4}$$

with the $L \times 1$ vector $\mathbf{h}_{i,j} = [h_{i,j}(0) \ h_{i,j}(1) \ \cdots \ h_{i,j}(L-1)]^T$ for $i = 1, 2, \ldots, n_t$ and $j = 1, 2, \ldots, n_r$. $\mathbf{E}$ is the $N \times n_r$ interference matrix that is composed of the thermal noise and jamming signals from the jammer.

The Bayesian estimator, defined in (3.16), of the channel vector is obtained as

$$\hat{\mathbf{h}} = [(\mathbf{I}_{n_r} \otimes \mathbf{S})^H \mathbf{Q}^{-1} (\mathbf{I}_{n_r} \otimes \mathbf{S}) + \mathbf{C}_H^{-1}]^{-1} (\mathbf{I}_{n_r} \otimes \mathbf{S})^H \mathbf{Q}^{-1} y, \tag{7.5}$$

where the channel covariance matrix $\mathbf{C}_H = \sigma^2 \mathbf{I}_{n_r n_t}$. In addition, the MSE of the Bayesian channel estimator, defined in (3.26), given by

$$\text{MSE} \ = \ \text{tr}[(\mathbf{I}_{n_r} \otimes \mathbf{S})^H \mathbf{Q}^{-1} (\mathbf{I}_{n_r} \otimes \mathbf{S}) + \mathbf{C}_H^{-1}]^{-1} \tag{7.6}$$

The objective is to find the training sequence set $\mathbf{S}$ in (7.2), which has block Toeplitz structure as we note from (7.3), that can minimize the MSE of the Bayesian channel estimator. Because of the structure constraint, it is not easy to find optimal sequences that satisfy the structure constraint. Thus, we may need to find sub-optimal sequences instead. In addition, approximating the block Toeplitz matrix $\mathbf{S}$ could be another way to tackle this problem.

REFERENCES

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol 1, no. 2, pp. 41–59, 1996.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[3] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "An overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vo2l. no. 3, pp. 281–302, Apr. 2003.

[4] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[5] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, pp. 452–460, Mar. 1999.

[6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[7] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804–1824, Jul. 2002.

[8] S. M. Alamouti, "A simple transimit diversity technique for wireless communications," *IEEE J. Slect. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[9] S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared error (LSSE) channel estimation," *IEE Proc. F.*, vol. 138, pp. 371–378, Aug 1991.

[10] G. Caire and U. Mitra, "Training sequence design for adaptive equalization of multi-user systems," *Thirty-Second Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1479–1483, Nov. 1998.

[11] C. Tellambura, Y. J. Guo, and S. K. Barton, "Channel estimation using aperiodic binary sequences," *IEEE Commun. Lett.*, vol. 2, pp. 140–142, May 1998.

[12] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform technique," *IEEE Trans. Commun.*, vol. 47, pp. 230–238, Feb. 1999.

[13] W. Chen and U. Mitra, "Training sequence optimization: comparison and an alternative criterion," *IEEE Trans. Commun.*, vol. 48, pp. 1987–1991, Dec. 2000.

[14] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. Int. Conf. Acoust, Speech, Signal Process.*, vol. 4, pp. 2105–2108, Salt Lake City, UT, May 7–11, 2001.

[15] W. H. Mow, *Sequence design for spread spectrum*, The Chinese University Press, Hong Kong, 1995.

[16] W. H. Mow, "A new unified construction of perfect root-of-unity sequences, " *IEEE Fourth International Symposium on Spread Spectrum Techniques and Applications (ISSSTA'96)*, vol. 3, pp. 955–959, Sep. 1996.

[17] D. C. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 531–532, July 1972.

[18] T. L. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," *37th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 22–24, 1999.

[19] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time coding modem for high data rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.

[20] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, Aug. 2000. Submitted for publication.

[21] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, pp. 384–391, Mar. 2003.

[22] B. Park and T. F. Wong, "Training sequence optimization in MIMO systems with colored noise," in *Proc. IEEE MILCOM '03*, Boston, MA, Oct. 2003.

[23] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. on Wireless Commun.*, submitted for publication, 2004

[24] E. G. Larsson and P. Stoica, *Space-time block coding for wireless communications*, Cambridge, UK: Cambridge University Press, 2003.

[25] S. M. Kay, *Fundamental of statistical signal processing: Estimation theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[26] J. M. Mendel, *Lessons in estimation theory for signal processing, communications, and control*, Englewood Cliffs, NJ: Prentice-Hall, 1995.

[27] R. M. Gray. (March/9/2004), *Toeplitz and circulant matrices: A review*, Revised Aug 2002. [Online]. Available: http://www-ee.stanford.edu/~gray/toeplitz.pdf

[28] Y. Song and S. D. Blostein, "Data detection in MIMO systems with co-channel interference," *IEEE 56th Vehicular Technology Conference*, vol. 1, pp. 3–7, Fall 2002.

[29] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*, Cambridge University Press, 1991.

[30] S. M. Kay, *Modern spectral estimation: Theory & application*, Prentice-Hall, New Jersey, 1988.

[31] S. L. Campbell and C. D. Meyer, *Generalized inverses of linear transformations*, Pitman, London, 1979.

[32] A. W. Marshall and I. Olkin, *Inequalities: Theory of majorization and its applications*, Academic Press, New York, 1968.

[33] E. K. P. Chong and S. H. Zak, *An introduction to optimization*, Wiley, New York, 1996.

[34] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[35] T. F. Wong and T. M. Lok, "Transmitter adaptation in multicode DS-CDMA systems," *IEEE Journal on Selected Areas in Commun.*, vol. 19, pp. 69–82, Jan. 2001.

[36] U. Grenander and G. Szego, *Toeplitz forms and their applications*, 2nd ed., Chelsea, New York, 1984.

[37] S. Haykin, *Adaptive filter theory*, 4th ed., Prentice-Hall, Upper Saddle River, NJ, 2002.

[38] J. G. Proakis, *Digital communications*, 4th ed., McGraw Hill, New York, 2001.

BIOGRAPHICAL SKETCH

Beomjin Park was born in Seoul, Korea, in 1969. He received his bachelor's degree in electronic engineering from Hankuk Aviation University, Korea, in 1995 and the Master of Science degree in electrical engineering from the University of Southern California in 1999. From 1995 to 1996, he was with Samsung Electronics Co., LTD., Korea, as a product engineer. Since January 2000, he has been pursing his Doctor of Philosophy degree in electrical and computer engineering at the University of Florida in the area of wireless communications. His research interests include space-time processing and channel estimation in multiple-input multiple-output (MIMO) systems.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Tan F. Wong, Chair
Assistant Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Yuguang "Michael" Fang,
Associate Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John M. Shea,
Assistant Professor of Electrical and Computer Engineering

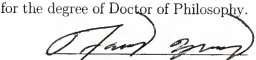
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Louis N. Cattafesta III,
Associate Professor of Mechanical and Aerospace Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

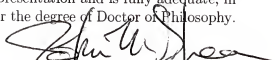
Tan F. Wong, Chair
Assistant Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



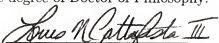
Yuguang "Michael" Fang,
Associate Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



John M. Shea,
Assistant Professor of Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Louis N. Cattafesta III,
Associate Professor of Mechanical and Aerospace Engineering

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 2004

Pramod Khargonekar

Pramod P. Khargonekar
Dean, College of Engineering

Kenneth J. Gerhardt
Interim Dean, Graduate School